


```
/clone_lib="CHORI-501"
1. .7523
/note="assembly_fragment:01480
fragment_chain:1
clone_end:T7
vector_side:left"
7624. .41360
/note="assembly_fragment:01718
fragment_chain:1"
41461. .47515
/note="assembly_fragment:00204
fragment_chain:1"
47616. .52829
/note="assembly_fragment:02002
fragment_chain:1"
52930. .90493
/note="assembly_fragment:00959
fragment_chain:1"
90594. .109805
/note="assembly_fragment:01522
fragment_chain:1"
109906. .136657
/note="assembly_fragment:01742
fragment_chain:1"
33294 a 32671 c 32974 g 37117 t 601 others
ORIGIN

Query Match 78.7%; Score 295; DB 2: Length 136657;
Best Local Similarity 100.0%; Pred. No. 1.3e-160; Indels 0; Gaps 0;
Matches 295; Conservative 0; Mismatches 0;

Qy 1 agggcccccgcacccctacctggctcccgcccccctctctccacccctcccgagcccccctaa 60
Db 54587 AGGCCCCCGCACCTCATCTGGCTCCGCCCTCTCTCCACCTCCGGACCCCTAA 54528

Qy 61 agggcgccggggcccaagccgagggcgctgocctgaccccgagcggaagggcccccagt 120
Db 54527 AGGGCGCGGGGGCCAAAGCCGAGCGCGCTGCGCTGACCCGAGCGGAAGGGCCCCAGT 54468

Qy 121 ctaggctcaatgcgggtggcgctctctttgacagcgcgctttgggacacacagcggg 180
Db 54467 CTAGTCTCTAATGCGGGTGGCGCTCTCTTTGACAGCGCGCGTGGGGACACACGCGGG 54408

Qy 181 acgagagataaggtgacataccagacagatttggtgcgcgctgatactctctcccg 240
Db 54407 ACGAGAGATAAGGTGACATACCAGACAGATTGGTGGCGCGCTGATCTCTCCG 54348

Qy 241 acaggaacacggagcgtatttaaagaccctatcgattactttacttctctgga 295
Db 54347 ACAGAAACGGGGAGCTATTAAAGACCCATCGATTACTTTATCTTTCTCTGGA 54293

RESULT 2
AC008412 AC008412 137545 bp DNA linear HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 5 clone CTC-281H14, WORKING DRAFT SEQUENCE,
DEFINITION 21 ordered pieces.
ACCESSION AC008412
VERSION AC008412.5 GI:9255970
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137545)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137545)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7708838.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14
-----
Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133998 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Estimated insert size: 136595; sum-of-contigs estimation
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2162: contig of 2162 bp in length
* 2163 2262: gap of unknown length
* 2263 5223: contig of 2961 bp in length
* 5224 5323: gap of unknown length
* 5324 15199: contig of 9876 bp in length
* 15200 15299: gap of unknown length
* 22312 22312: contig of 7013 bp in length
* 22313 22412: gap of unknown length
* 22413 37107: contig of 14695 bp in length
* 37108 37207: gap of unknown length
* 37208 43863: contig of 6656 bp in length
* 43864 43963: gap of unknown length
* 43964 46965: contig of 3002 bp in length
* 46966 47065: gap of unknown length
* 47066 49477: contig of 2412 bp in length
* 49478 49577: gap of unknown length
* 49578 54523: contig of 4946 bp in length
* 54524 54624: gap of unknown length
* 54624 68414: contig of 13791 bp in length
* 68415 68514: gap of unknown length
* 68515 72851: contig of 4336 bp in length
* 72851 72951: gap of unknown length
* 72951 88213: contig of 15263 bp in length
* 88214 88313: gap of unknown length
* 88314 98538: contig of 10225 bp in length
* 98539 98639: gap of unknown length
* 98639 107774: contig of 9136 bp in length
* 107775 107874: gap of unknown length
* 107875 115145: contig of 7271 bp in length
* 115146 115245: gap of unknown length
* 115246 118552: contig of 3307 bp in length
* 118553 118652: gap of unknown length
* 118653 128198: contig of 9546 bp in length
* 128199 128298: gap of unknown length
* 128299 130123: contig of 1825 bp in length
* 130124 130223: gap of unknown length
* 130224 134038: contig of 3815 bp in length
* 134039 134138: gap of unknown length
* 134139 136470: contig of 2332 bp in length
* 136471 137545: gap of unknown length
* 136571 137545: contig of 975 bp in length.
Location/Qualifiers
1. .137545
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```


ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 79364)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oshan,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,W.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S.,
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabar,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,K., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

TITLE
JOURNAL

2 (bases 1 to 79364)

Worley,K.C.

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062509.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHEJ
Center clone name: CH230-53024
----- Summary Statistics
findPhrapList
Consensus quality: 57802 bases at least Q40
Consensus quality: 64028 bases at least Q30
Consensus quality: 67953 bases at least Q20
Estimated insert size: 35926; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
2616: contig of 2616 bp in length
2617: gap of unknown length
2717: contig of 3060 bp in length
5876: gap of unknown length
5877: contig of 1817 bp in length
7693: gap of unknown length
7694: contig of 1191 bp in length
8984: gap of unknown length
8985: contig of 1963 bp in length
9085: gap of unknown length
11047: contig of 1302 bp in length
11148: gap of unknown length
12449: contig of 1070 bp in length
12549: gap of unknown length
13619: contig of 1070 bp in length
13620: gap of unknown length
13719: contig of 2198 bp in length
15917: gap of unknown length
16017: contig of 2042 bp in length
18059: gap of unknown length
18159: gap of unknown length
19630: contig of 1471 bp in length
19631: gap of unknown length
19730: contig of 1550 bp in length
21280: gap of unknown length
21380: contig of 1341 bp in length
22721: gap of unknown length
22821: contig of 1327 bp in length
24148: gap of unknown length
24248: contig of 1785 bp in length
26033: gap of unknown length
26133: contig of 1388 bp in length
27521: gap of unknown length
27621: contig of 1949 bp in length
29570: gap of unknown length
29670: contig of 1334 bp in length
31004: gap of unknown length
31104: contig of 1463 bp in length
32567: gap of unknown length
32667: contig of 1676 bp in length
34343: gap of unknown length
34443: contig of 1203 bp in length
35645: gap of unknown length
35746: contig of 1082 bp in length
36828: gap of unknown length
36928: contig of 1774 bp in length
38702: gap of unknown length
38802: contig of 1467 bp in length
40269: gap of unknown length
40369: contig of 1269 bp in length
41638: gap of unknown length
41738: contig of 1283 bp in length
43021: gap of unknown length
43121: contig of 1282 bp in length
44403: gap of unknown length
44503: contig of 1293 bp in length
45796: gap of unknown length
45896: contig of 1263 bp in length
47159: gap of unknown length
47259: contig of 1405 bp in length
48664: gap of unknown length
48764: contig of 1507 bp in length
50271: gap of unknown length
50371: contig of 1386 bp in length
51575: gap of unknown length
51758: contig of 1653 bp in length
53511: gap of unknown length
53610: contig of 1611 bp in length
55221: gap of unknown length


```
Qy 361 ctgcacagcgacgc 375
    |||||
Db 361 ctgcacagcgacgc 375

RESULT 2
AAS09960
ID AAS09960 standard; DNA; 1072 BP.
XX
AC AAS09960;
XX
DT 24-OCT-2001 (first entry)
XX
DE DNA encoding mCsx/Nkx2.5 homology domains A1 + A2, and intervening DNA.
XX
KW mCsx/Nkx2.5 homology domain; cardiac enhancer; cardiac cell;
KW cardiomyocyte induction; therapeutic; heart tissue; gene therapy;
KW mouse; ds.
XX
OS Mus musculus.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX
PR 14-JAN-2000; 2000US-0176419.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lee IW, Izumo S;
XX
WPI; 2001-451809/48.
XX
PT New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
PS Claim 1; Fig 5B; 66pp; English.
XX
CC The sequence represents the coding sequence of cardiac enhancer
CC mCsx/Nkx2.5 homology domains A1 + A2 and intervening sequence. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
SQ Sequence 1072 BP; 273 A; 290 C; 297 G; 212 T; 0 other;

Query Match 100.0%; Score 375; DB 22; Length 1072;
Best Local Similarity 100.0%; Pred. No. 9.3e-184;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agccccccgacccctcctgctccgcgcgccttctctccacccctcccgagccctaa 60
    |||||
Db 1 agccccccgacccctcctgctccgcgcgccttctctccacccctcccgagccctaa 60

Qy 61 agggcgcgggggcccaagcgaggcgctgcctgaccccgagcggaagggccccagt 120
    |||||
Db 61 agggcgcgggggcccaagcgaggcgctgcctgaccccgagcggaagggccccagt 120

Qy 121 ctaggtctaatgcgggtgcgtctcctttgacaggcgcggtttggggacaacagcgggg 180
    |||||
Db 121 ctaggtctaatgcgggtgcgtctcctttgacaggcgcggtttggggacaacagcgggg 180

Qy 181 acgagagataaggtgacataccagacagatttgctgcgcgctgatactctctccg 240
    |||||
Db 181 acgagagataaggtgacataccagacagatttgctgcgcgctgatactctctccg 240
```

```
Qy 241 acaggaacgcggagctatttaaaagacccctatcgattacttattcttctggaagct 300
    |||||
Db 241 acaggaacgcggagctatttaaaagacccctatcgattacttattcttctggaagct 300

Qy 301 tcttcggagagacaaaaagatgttccctgcctaaagacacaaagggccacacacggagggt 360
    |||||
Db 301 tcttcggagagacaaaaagatgttccctgcctaaagacacaaagggccacacacggagggt 360

Qy 361 ctgcacagcgacgc 375
    |||||
Db 361 ctgcacagcgacgc 375

RESULT 3
AAS09961
ID AAS09961 standard; DNA; 7836 BP.
XX
AC AAS09961;
XX
DT 24-OCT-2001 (first entry)
XX
DE Genomic DNA #1 encoding human Csx/Nkx2.5.
XX
KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
KW therapeutic; heart tissue; gene therapy; human; ds.
XX
OS Homo sapiens.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX
PR 14-JAN-2000; 2000US-0176419.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lee IW, Izumo S;
XX
WPI; 2001-451809/48.
XX
PT New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
PS Disclosure; Fig 4A; 66pp; English.
XX
CC The sequence represents the genomic sequence #1 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
SQ Sequence 7836 BP; 2164 A; 1938 C; 1907 G; 1825 T; 2 other;

Query Match 100.0%; Score 375; DB 22; Length 7836;
Best Local Similarity 100.0%; Pred. No. 8.9e-184;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agccccccgacccctcctgctccgcgcgccttctctccacccctcccgagccctaa 60
    |||||
Db 3588 agggcccccgacccctcctgctccgcgcgccttctctccacccctcccgagccctaa 3647

Qy 61 agggcgcgggggcccaagcgaggcgctgcctgaccccgagcggaagggccccagt 120
    |||||
Db 3648 agggcgcgggggcccaagcgaggcgctgcctgaccccgagcggaagggccccagt 3707

Qy 121 ctaggtctaatgcgggtgcgtctcctttgacaggcgcggtttggggacaacagcgggg 180
```

[illegible]

Search completed: July 5, 2002, 06:17:00
Job time: 5398 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 04:46:17 : Search time 80.62 Seconds
(without alignments)
1142.552 Million cell updates/sec

Title: US-09-761-466-1
Perfect score: 375
Sequence: 1 aggcaccccgaccctcacc.....agggtctcacacaggcaacgc 375

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 6

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cqn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cqn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cqn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cqn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cqn2_6/ptodata/2/ina/backfiles.seq.*

Fred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB	ID	Description
.....						

No matches found

Search completed: July 5, 2002, 06:10:55
Job time: 5078 sec

GenBank version 4.4
Copyright (c) 1994 - 2000 CompuLink Ltd.

AM nucleic - nucleic search, using sw model

Pub on: July 5, 2002, 04:51:53 : Search time 448.65 Seconds
(without alignments)
1807.467 Million cell updates/sec

Title: US-09-761-466-1
Perfect score: 375
Sequence: 1 aagcccccacacccatc.....agatttcacacagcagc 375

Scoring table: G1G0_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2197556 seqs, 1081744927 residues

Word size : 30

Total number of hits satisfying chosen parameters: 7

Minimum hit seq length: 6

Maximum hit seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Pending Patents, NA, Main: *
- 1: /cqn2_6/ptodata/2/pna/US06_COMB.seq.*
 - 2: /cqn2_6/ptodata/2/pna/US06_COMB.seq.*
 - 3: /cqn2_6/ptodata/2/pna/US07_COMB.seq.*
 - 4: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 5: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 6: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 7: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 8: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 9: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 10: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 11: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 12: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 13: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 14: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 15: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 16: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 17: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 18: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 19: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 20: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 21: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 22: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 23: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 24: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 25: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 26: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 27: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 28: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 29: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 30: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 31: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 32: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 33: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 34: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 35: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 36: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 37: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 38: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 39: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 40: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 41: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 42: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*
 - 43: /cqn2_6/ptodata/2/pna/US08_COMB.seq.*

44:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
45:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
46:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
47:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
48:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
49:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
50:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
51:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
52:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
53:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
54:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
55:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
56:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
57:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
58:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
59:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
60:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
61:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
62:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
63:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
64:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
65:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
66:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
67:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
68:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
69:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
70:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
71:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
72:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
73:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
74:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11
75:	/cqn2_6/ptodata/2/pna/US08_COMB.seq.*	MR, Seq. 11

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	JB	JB	Reset	Item
1	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1
2	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1
3	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1
4	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1
5	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1
6	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1
7	US01-01511-1	375	100.0	375	1	1	1	Sequence 1, Appl. 1

ALL INVENTS

RESULT 1
PCI-US01-01511-1
: GENERAL INFORMATION:
: APPLICANT: Both Israel Biocore Medical Center
: TITLE OF INVENTION: Cardiac-Cell Specific Embryonic Elements
: TITLE OF INVENTION: Cardiac-Cell Specific Embryonic Elements
: FILE REFERENCE: 01944/06962
: CURRENT APPLICATION NUMBER: PCI-US-1/1511
: PRIOR APPLICATION NUMBER: 01944/06962
: PRIOR APPLICATION NUMBER: 01944/06962
: NUMBER OF SEQ ID NOS: 20
: SEQUENCE: FastSeq for Windows Version 4.1
: SEQ ID NO: 1
: LENGTH: 375
: TYPE: DNA
: ORGANISM: Mus musculus
PCI-US01-01511-1

```
Query Match      100.0%; Score 375; DB 1; Length 375;
Best Local Similarity 100.0%; Pred. No. 2.4e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agggccccgcacccctcactcgtgctcccgccctcccttctctccaccctcccggaaccctctaa 60
Db 1 agggccccgcacccctcactcgtgctcccgcccttctctccaccctcccggaaccctctaa 60

Qy 61 agggcgcgcgggcccaagcgaggcgctgcgctgaccccgagcggaagggccccagt 120
Db 61 agggcgcgcgggcccaagcgaggcgctgcgctgaccccgagcggaagggccccagt 120

Qy 121 ctaggctctaagcggggtggcgctcctctttgacagcgcggtttggggacacacgagggg 180
Db 121 ctaggctctaagcggggtggcgctcctctttgacagcgcggtttggggacacacgagggg 180

Qy 181 acgagagataaggtgacataccagagcagatttggcgcgcgctgatactctctcccg 240
Db 181 acgagagataaggtgacataccagagcagatttggcgcgcgctgatactctctcccg 240

Qy 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttcttgaaagct 300
Db 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttcttgaaagct 300

Qy 301 tcttgcggagagacaaaagatgttccctgcctaaagacacaaagccacacacgaggggt 360
Db 301 tcttgcggagagacaaaagatgttccctgcctaaagacacaaagccacacacgaggggt 360

Qy 361 ctgcacagggcgagcg 375
Db 361 ctgcacagggcgagcg 375

RESULT 2
US-09-761-466-1
; Sequence 1, Application US/09761466
; GENERAL INFORMATION:
; APPLICANT: Lee, Ike W.
; APPLICANT: Izumo, Seigo
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: US/09/761.466
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-761-466-1
```

```
Query Match      100.0%; Score 375; DB 30; Length 375;
Best Local Similarity 100.0%; Pred. No. 2.4e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agggccccgcacccctcactcgtgctcccgcccttctctccaccctcccggaaccctctaa 60
Db 1 agggccccgcacccctcactcgtgctcccgcccttctctccaccctcccggaaccctctaa 60

Qy 61 agggcgcgcgggcccaagcgaggcgctgcgctgaccccgagcggaagggccccagt 120
Db 61 agggcgcgcgggcccaagcgaggcgctgcgctgaccccgagcggaagggccccagt 120

Qy 121 ctaggctctaagcggggtggcgctcctctttgacagcgcggtttggggacacacgagggg 180
Db 121 ctaggctctaagcggggtggcgctcctctttgacagcgcggtttggggacacacgagggg 180
```

```
Qy 181 acgagagataaggtgacataccagagcagatttggcgcgcgctgatactctctcccg 240
Db 181 acgagagataaggtgacataccagagcagatttggcgcgcgctgatactctctcccg 240

Qy 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttcttgaaagct 300
Db 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttcttgaaagct 300

Qy 301 tcttgcggagagacaaaagatgttccctgcctaaagacacaaagccacacacgaggggt 360
Db 301 tcttgcggagagacaaaagatgttccctgcctaaagacacaaagccacacacgaggggt 360

Qy 361 ctgcacagggcgagcg 375
Db 361 ctgcacagggcgagcg 375

RESULT 3
PCT-US01-01511-3
; Sequence 3, Application PC/TUS0101511
; GENERAL INFORMATION:
; APPLICANT: Beth Israel Deaconess Medical Center
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 01948/069W02
; CURRENT APPLICATION NUMBER: PCT/US01/01511
; CURRENT FILING DATE: 2000-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01-01511-3
```

```
Query Match      100.0%; Score 375; DB 1; Length 1072;
Best Local Similarity 100.0%; Pred. No. 2.5e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agggccccgcacccctcactcgtgctcccgcccttctctccaccctcccggaaccctctaa 60
Db 1 agggccccgcacccctcactcgtgctcccgcccttctctccaccctcccggaaccctctaa 60

Qy 61 agggcgcgcgggcccaagcgaggcgctgcgctgaccccgagcggaagggccccagt 120
Db 61 agggcgcgcgggcccaagcgaggcgctgcgctgaccccgagcggaagggccccagt 120

Qy 121 ctaggctctaagcggggtggcgctcctctttgacagcgcggtttggggacacacgagggg 180
Db 121 ctaggctctaagcggggtggcgctcctctttgacagcgcggtttggggacacacgagggg 180

Qy 181 acgagagataaggtgacataccagagcagatttggcgcgcgctgatactctctcccg 240
Db 181 acgagagataaggtgacataccagagcagatttggcgcgcgctgatactctctcccg 240

Qy 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttcttgaaagct 300
Db 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttcttgaaagct 300

Qy 301 tcttgcggagagacaaaagatgttccctgcctaaagacacaaagccacacacgaggggt 360
Db 301 tcttgcggagagacaaaagatgttccctgcctaaagacacaaagccacacacgaggggt 360

Qy 361 ctgcacagggcgagcg 375
Db 361 ctgcacagggcgagcg 375
```

```
RESULT 4
US-09-761-466-3
```

```

; Sequence 3, Application US/09761465
; GENERAL INFORMATION:
; APPLICANT: Lee, Ike W.
; APPLICANT: Izumo, Seigo
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: US/09/761,466
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-761-466-3

Query Match 100.0%; Score 375; DB 30; Length 1072;
Best Local Similarity 100.0%; Prod. No. 2,5e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagccccgcacccctcatcttgctccgcgcctctctctccacccctccgcagacccctaa 60
DB 1 aagccccgcacccctcatcttgctccgcgcctctctctccacccctccgcagacccctaa 60

QY 61 aagggcgagggcccaacccgagagcctagcctcagcctcagcccaagcagaccccaat 120
DB 61 aagggcgagggcccaacccgagagcctagcctcagcctcagcccaagcagaccccaat 120

QY 121 ctatgctcttaagcggtgagctatcagcagacacatctctctcagcagcagctctccag 180
DB 121 ctatgctcttaagcggtgagctatcagcagacacatctctctcagcagcagctctccag 180

QY 301 tcttgagagagacacacacacacacacacacacacacacacacacacacacacacacac 360
DB 301 tcttgagagagacacacacacacacacacacacacacacacacacacacacacacacac 360

QY 361 ctgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 375
DB 361 ctgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 375

RESULT 5
US-09-761-466-4
; Sequence 4, Application US/09761466
; GENERAL INFORMATION:
; APPLICANT: Lee, Ike W.
; APPLICANT: Izumo, Seigo
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: US/09/761,466
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(7838)
; OTHER INFORMATION: n = A,T,C or G
US-09-761-466-4

Query Match 100.0%; Score 375; DB 30; Length 7838;
Best Local Similarity 100.0%; Prod. No. 2,6e-194;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagccccgcacccctcatcttgctccgcgcctctctctccacccctccgcagacccctaa 60
DB 1 aagccccgcacccctcatcttgctccgcgcctctctctccacccctccgcagacccctaa 60

QY 61 aagggcgagggcccaacccgagagcctagcctcagcctcagcccaagcagaccccaat 120
DB 61 aagggcgagggcccaacccgagagcctagcctcagcctcagcccaagcagaccccaat 120

QY 121 ctatgctcttaagcggtgagctatcagcagacacatctctctcagcagcagctctccag 180
DB 121 ctatgctcttaagcggtgagctatcagcagacacatctctctcagcagcagctctccag 180

QY 301 tcttgagagagacacacacacacacacacacacacacacacacacacacacacacacac 360
DB 301 tcttgagagagacacacacacacacacacacacacacacacacacacacacacacacac 360

QY 361 ctgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 375
DB 361 ctgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 375

RESULT 5
PCT-US01-01511-4
; Sequence 4, Application PCT/US0101511
; GENERAL INFORMATION:
; APPLICANT: Both Israel Deaconess Medical Center
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: PCT/US01/01511
; PRIOR FILING DATE: 2000-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature

```

```
Db 3648 aggggaggggggcaagcggagggcgctgcgctgaccccgagcggaaggggcccaagt 3707
QY 121 ctaggctcctaagcggtggtgctctctttgacagcgcggtttgaggacacacagcgggg 180
Db 3708 ctaggctcctaagcggtggtgctctctttgacagcgcggtttgaggacacacagcgggg 3767
QY 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactcctctccc 240
Db 3768 acgagagataaggtgacataccagagcagatttggcgcgctgatactcctctccc 3827
QY 241 acaggaaacggagcgctatttaaaagaccctatcgatttatttcttctgaaaagct 300
Db 3828 acaggaaacggagcgctatttaaaagaccctatcgatttatttcttctgaaaagct 3887
QY 301 tcttgaggagagacaaagatgttcctgcgttaaaagacacacaaaggcgaggggt 360
Db 3888 tcttgaggagagacaaagatgttcctgcgttaaaagacacacaaaggcgaggggt 3947
QY 361 ctgcacagggcgacgc 375
Db 3948 ctgcacagggcgacgc 3962
```

```
RESULT 7
US-60-182-316-1254
; Sequence 1254, Application US/60182316
; GENERAL INFORMATION:
; APPLICANT: Curtis, Anne L.
; APPLICANT: Lagace, Robert E.
; APPLICANT: Klingler, Tod M.
; APPLICANT: Stuve, Laura L.
; FILE REFERENCE: PX-0003 P
; CURRENT APPLICATION NUMBER: US/60/182,316
; CURRENT FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 14,630
; SOFTWARE: PERL Program
; SEQ ID NO 1254
; LENGTH: 338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: CpG_991027_B15_masked_fa.Contig29958
US-60-182-316-1254
```

```
Query Match 21.3%; Score 80; DB 57; Length 338;
Best Local Similarity 100.0%; Pred. No. 2.3e-32;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 aagctcttcgagagacaaagatgttcctgcgttaaaagacacacaaaggcgacacgg 355
Db 10 aagctcttcgagagacaaagatgttcctgcgttaaaagacacacaaaggcgacacgg 69

QY 356 aggggtctgcacagggcgacgc 375
Db 70 aggggtctgcacagggcgacgc 89
```

Search completed: July 5, 2002, 07:32:11
Job time: 9618 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

GM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 05:31:03 : Search time 248.75 seconds
(without alignments)
2425.605 Million cell updates/sec

Title: us-09-761-466-1
Perfect score: 375
Sequence: 1 agggcccccgcacccctcgc.....agggctctgcacaggcgaagc 375

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1089956 seqs, 804492379 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents.NA.New.*
1: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---------------	-------	----------------	--------	-------	-------------

No matches found

Search completed: July 5, 2002, 07:36:39
Job time: 7415 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 04:44:12 : Search time 2730.19 Seconds
(without alignments)
1853.848 Million cell updates/sec

Title: US-09-761-466-1
Percent score: 375
Sequence: 1 aagaccccccacaccctcacc.....aaggctctcacaggcgaacac 375

Scoring table: Q100_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13735207 seqs, 6748477542 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
-----	-----	-----	-----	-----

No matches found

Search completed: July 5, 2002, 05:32:51
Job time: 2919 sec


```
misc_feature 1. 7523
/note="assembly_fragment:01480
fragment_chain:1
clone_end:T7
vector_side:left"
7624. 41360
/note="assembly_fragment:01718
fragment_chain:1"
41461. 47515
/note="assembly_fragment:00204
fragment_chain:1"
47616. 52829
/note="assembly_fragment:02002
fragment_chain:1"
52930. 90493
/note="assembly_fragment:00959
fragment_chain:1"
90594. 109805
/note="assembly_fragment:01522
fragment_chain:1"
109906. 136657
/note="assembly_fragment:01742
fragment_chain:1"
BASE COUNT 33294 a 32671 c 32974 g 37117 t 601 others
ORIGIN
```

```
Query Match 100.0%; Score 51; DB 2; Length 136657;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcctctttaaggcttgatgtcgaactgcatgtgtacacttaag 51
|||||
Db 53556 TGCTCCTTTAAGGCTTGAATGCTGCAACTGTCATGTGTACACTTAAG 53506
|||||
```

RESULT 2

```
AC008412
LOCUS AC008412 137545 bp DNA linear HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-281H14, WORKING DRAFT SEQUENCE,
21 ordered pieces.
ACCESSION AC008412
VERSION AC008412.5 GI:9255970
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137545)
DOE Joint Genome Institute.
Unpublished
Sequencing of Human Chromosome 5
2 (bases 1 to 137545)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7708838.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14
-----
```

```
Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133398 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Sum-of-contigs estimation
```

```
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
```

```
1 2162: contig of 2162 bp in length
2262: gap of unknown length
5223: contig of 2961 bp in length
5224 5223: gap of unknown length
5324 15199: contig of 9876 bp in length
15200 15299: gap of unknown length
22312: contig of 7013 bp in length
22313 22412: gap of unknown length
22413 37107: contig of 14695 bp in length
37108 37207: gap of unknown length
37208 43863: contig of 6656 bp in length
43864 43963: gap of unknown length
43964 46965: contig of 3002 bp in length
46966 47065: gap of unknown length
47066 49477: contig of 2412 bp in length
49478 49577: gap of unknown length
49578 54523: contig of 4946 bp in length
54524 54623: gap of unknown length
54624 68414: contig of 13791 bp in length
68415 68514: gap of unknown length
68515 72850: contig of 4336 bp in length
72851 88213: contig of 15263 bp in length
88214 88313: gap of unknown length
88314 98538: contig of 10225 bp in length
98539 98638: gap of unknown length
98639 107774: contig of 9136 bp in length
107775 107874: gap of unknown length
107875 115145: contig of 7271 bp in length
115146 115245: gap of unknown length
115246 118552: contig of 3307 bp in length
118553 118653: gap of unknown length
118654 128198: contig of 9546 bp in length
128199 128298: gap of unknown length
128299 130123: contig of 1825 bp in length
130124 130224: gap of unknown length
130224 134038: contig of 3815 bp in length
134039 134138: gap of unknown length
134139 136470: contig of 2332 bp in length
136471 137545: contig of unknown length
137546 137547: contig of 975 bp in length.
```

FEATURES

```
Location/Qualifiers
1..137545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-281H14"
/clone_lib="Caltech human BAC library C"
BASE COUNT 37277 a 32308 c 31466 g 34494 t 2000 others
ORIGIN
```

```
Query Match 100.0%; Score 51; DB 2; Length 137545;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 tgcctctttaaggcttgatgtcgaactgcatgtgtacacttaag 51
|||||
Db 29472 TGCTCCTTTAAGGCTTGAATGCTGCAACTGTCATGTGTACACTTAAG 29522
|||||
RESULT 3
```

```

AL669821      144702 bp   DNA   linear   HTG 31-JAN-2002
LOCUS        Homo sapiens chromosome 6 clone Xbac-116A1, *** SEQUENCING IN
DEFINITION   PROGRESS ***, 4 unordered pieces.
ACCESSION    AL669821
VERSION      AL669821.4 GI:18157151
KEYWORDS     HTG: H'GS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLT'P.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Mashreghi-Mohammadi.M.
Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk clone requests: clonerequest@sanger.ac.uk
On Jan 15, 2002 this sequence version replaced qi:18152649.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BPG116A1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 144176 bases at least Q40
Consensus quality: 144232 bases at least Q30
Consensus quality: 144299 bases at least Q20
Insert size: 144402; sum-of-contigs
Insert size: 167909; 14.3% error; agarose-fp
Quality coverage: 10.72x in Q20 bases; sum-of-contigs quality
coverage: 9.55x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* j 75928: contig of 75928 bp in length
* 75929 76028: gap of 100 bp
* 76029 116346: contig of 40318 bp in length
* 116347 116446: gap of 100 bp
* 116447 131258: contig of 14812 bp in length
* 131259 131358: gap of 100 bp
* 131359 144702: contig of 13344 bp in length.
*
* Location/Qualifiers
* 1..144702
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="6"
  /clone="XXbac-116A1"
  /clone_lib="CHGR1-501"
  1..75928
    /note="assembly_fragment:03023"
    fragment_chain:1
      clone_end:T7
      vector_side:left
  76029..116346
    /note="assembly_fragment:03493"
    fragment_chain:1
  116447..131258
    /note="assembly_fragment:03524"
    fragment_chain:1
  131359..144702
    /note="assembly_fragment:01454"
    fragment_chain:1

```


PD 19-JUL-2001.
XX
XX CC 16-JAN-2001; 2001WO-US01511.
XX PF
XX PR 14-JAN-2000; 2000US-0176419.
XX XX
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI
XX PI Lee IW, Izumo S;
XX XX
XX WPI; 2001-451809/48.
XX
XX PT New cardiac specific cell enhancer elements, useful for specifically
XX expressing gene in cardiac cell, as earlier marker of cardiomyocyte
XX induction, e.g. for optimizing cardiomyocyte induction -
XX
XX PS Claim 1; Fig 5B; 66pp; English.
XX
XX CC The sequence represents the coding sequence of cardiac enhancer
XX mCsx/Nkx2.5 homology domains A1 + A2 and intervening sequence. The
XX nucleic acid is useful for specifically expressing a gene in a cardiac
XX cell, as an earlier marker of cardiomyocyte induction, e.g. for
XX cell optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
XX specific manner are useful for the targeted expression of genes encoding
XX therapeutic proteins for the treatment of damaged heart tissue. Cardiac
XX specific enhancer elements may be used for gene therapy.
XX
XX SQ Sequence 1072 BP; 273 A; 290 C; 297 G; 212 T; 0 other;

Query Match 100.0%; Score 51; DB 22; Length 1072;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcctctttaagggttgatgtctgcaactgtcatgtgtacacttaaaag 51
|||||
Db 1022 tgcctctttaagggttgatgtctgcaactgtcatgtgtacacttaaaag 1072

RESULT 3
AAS09961
ID AAS09961 standard; DNA; 7836 BP.
XX
XX AC AAS09961;
XX DT
XX DE 24-OCT-2001 (first entry)
XX Genomic DNA #1 encoding human Csx/Nkx2.5.
XX
XX KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
XX KW therapeutic; heart tissue; gene therapy; human; ds.
XX OS Homo sapiens.
XX PN WO200151006-A2.
XX PD 19-JUL-2001.
XX
XX PF 16-JAN-2001; 2001WO-US01511.
XX XX
XX PR 14-JAN-2000; 2000US-0176419.
XX
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI
XX PI Lee IW, Izumo S;
XX XX
XX WPI; 2001-451809/48.
XX
XX PT New cardiac specific cell enhancer elements, useful for specifically
XX expressing gene in cardiac cell, as earlier marker of cardiomyocyte
XX induction, e.g. for optimizing cardiomyocyte induction -
XX
XX PS Disclosure; Fig 4A; 66pp; English.

XX

CC The sequence represents the genomic sequence #1 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
XX SQ Sequence 7836 BP; 2164 A; 1938 C; 1907 G; 1825 T; 2 other;

Query Match 100.0%; Score 51; DB 22; Length 7836;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcctctttaagggttgatgtctgcaactgtcatgtgtacacttaaaag 51
|||||
Db 4609 tgcctctttaagggttgatgtctgcaactgtcatgtgtacacttaaaag 4659

Search completed: July 5, 2002, 06:17:00
Job time: 5398 sec

SeqWare version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

M nucleotide - nucleotide search, using sw model

Run on: July 5, 2002, 06:10:55 - Search time 80.62 seconds
(without alignments)
155,487 Million cell updates/sec

Title: US-09-761-466-2
Perfect score: 51
Sequence: 1 tgcctctttaaaqcttga.....tgctatgtatataacttaaaa 51

Scoring table: GLCO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum hit seq length: 0
Maximum hit seq length: 200000000

Post-processing: listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cdu2_5/ptodata/2/ina/5A_0.MB.seq.*
2: /cdu2_5/ptodata/2/ina/5B_0.MB.seq.*
3: /cdu2_5/ptodata/2/ina/5A_0.MB.seq.*
4: /cdu2_5/ptodata/2/ina/5B_0.MB.seq.*
5: /cdu2_5/ptodata/2/ina/pt015_0.MB.seq.*
6: /cdu2_5/ptodata/2/ina/backfiles1.seq.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

80 matches found

Search completed: July 5, 2002, 06:10:55
Job time: 5076 sec

Genware Vers. 3.6.4.5
Copyright (c) 1995 - 2000 CompuGen Ltd.

EM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 05:42:51 : Search time 2730.19 Seconds
(without alignments)
252.123 Million cell updates/sec

Title: us-09-761-466-2
Percent score: 51
Sequence: 1 tttctcttttaaacqggttaa.....ttgtcatqctacacttaaa 51

Scoring table: 0.170,NDC
Gapop 60.0 , Gapext 60.0

Searched: 14736207 seqs, 6748477542 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : EST:
1: em_estha:
2: em_esthm:
3: em_estim:
4: em_estma:
5: em_estov:
6: em_estpl:
7: em_estro:
8: em_hic:
9: qb_est1:
10: qb_est2:
11: qb_hic:
12: qb_oss:
13: em_oss_him:
14: em_oss_hov:
15: em_oss_kib:
16: em_oss_vrt:

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Prod. No.	Score	Match	Length	DB	ID	Description
.....

87 matches found

Search completed: July 5, 2002, 05:42:51
Job time: 2919 sec


```

Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bpg250J14
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 134624 bases at least Q40
Consensus quality: 135417 bases at least Q30
Consensus quality: 135772 bases at least Q20
Insert size: 136057; sum-of-contigs
Insert size: 15180; 6.1% error; agarose-fp
Quality coverage: 7.81x in Q20 bases; sum-of-contigs Quality
coverage: 9.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
* 7524 7623: contig of 7523 bp in length
* 7624 7623: gap of 100 bp
* 7624 41360: contig of 33737 bp in length
* 41361 41460: gap of 100 bp
* 41461 47515: contig of 6055 bp in length
* 47516 47615: gap of 100 bp
* 47616 52829: contig of 5214 bp in length
* 52830 52929: gap of 100 bp
* 52930 90493: contig of 37564 bp in length
* 90494 90593: gap of 100 bp
* 90594 109805: contig of 19212 bp in length
* 109806 109905: gap of 100 bp
* 109906 136657: contig of 26752 bp in length.
FEATURES
Source
1..136657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXBac-250J14"
/clone_lib="CHORI-501"
1..7523
/note="assembly_fragment:01480
fragment_chain:1
clone_end:77
vector_side:left"
7624..41360
/note="assembly_fragment:01718
fragment_chain:1"
41461..47515
/note="assembly_fragment:00204
fragment_chain:1"
47616..52829
/note="assembly_fragment:02002
fragment_chain:1"
52930..90493
/note="assembly_fragment:00959
fragment_chain:1"
90594..109805
/note="assembly_fragment:01522
fragment_chain:1"
109906..136657
/note="assembly_fragment:01742
fragment_chain:1"
BASE COUNT 33294 a 32671 c 32974 g 37117 t 601 others
ORIGIN

```

QY	1	aggccccccggaacctatctatctgctccggcccccttctctctcaacctccccggaacctctaa	60
Db	54587	AGGCCCCCGGACCCCTCATCTCTGGCTCCGGCCCTCTCTCTCACCTCCCGAGACCCCTTAA	54528
QY	61	aggggcgcgggggcccaagccgagggcgctcgctgaccccccgagcgaaagggccccaggt	120
Db	54527	AGGGCGCGGGGGCCCAAGCCAGGGGCGCTGCGCCTGACCCCGAGCGGAAGGGGCCCCAGT	54468
QY	121	ctaggtcctaattgcgggtggtgcgtccctttgacagggcggtttgagggacaacaagcgaggg	180
Db	54467	CTAGGTCCTATGCGGTGGCGTCTCCTTTGACAGCGCGGCTTTGGGGACAACACGCGGG	54408
QY	181	acgagagataaggtgacataccagagcagatttggtgcgcgcgtgatactctctctccg	240
Db	54407	ACGAGAGATAAGGTGACATACACAGAGCAGATTGGTGCGCGCGTGATCTCTCTCTCCG	54348
QY	241	acaggaacgcggagctatttaagaagccccctatcgattacttlatcttcttct--ggaaa	297
Db	54347	ACAGGAAACGGCGAGCTATTTAAAGACCCCTATCGATTACTTTATCTTCTCTGGAGGAAA	54288
QY	298	gctcttcgcgagagacaaaagatgttccctgcctctaaagacacaaagggccacacaacgag	357
Db	54287	GCTCTTCGGAGAGACAAAAGATGTTCCCTGCTCTAAAGACAACAAGGCCACACAACGGAG	54228
QY	358	ggctgcagagcgacgcacaattcgcgcggggaaagcaaaaaacacactgacgcttaga	417
Db	54227	GGTCTGCACAGCGACGCACAATTCGGCGCGGGGAAAGCAAAAACACACTGACGCTTAGA	54168
QY	418	gtgcacaaactgtgtgtctccagagcagctccagagtcgcggagggagcgctggggggcgg	477
Db	54167	GTGCACAAACGTGTGTGTCTCCACAGCAGCTCCACAGTGCAGGTCGCGCACGGGACGCTGGGGCGG	54108
QY	478	cgagggcacccacagtatgtctctgtgccttggaaagtgttttttaccgtatcgcg	537
Db	54107	CGAGGGGCACCCACAGTATGTGTTCTGTGCCCCTTGGAAGAGTTTTTTTTCACCGTATGCG	54048
QY	538	cgtaaaacacgcacacacagagaaagtgcactgtgcacttagggcgccctgtgtgtaccgt	597
Db	54047	CGTAAACACGCACACACAGAGAAAGTGACTGTGCACTTAGGGCGCCTGTGTGTACCCGT	53988
QY	598	gtcgttttagcgaatttaagcacatacaggcgcgggcccatggtctacgcgctgtgtaatccc	657
Db	53987	GTCTGTTTTAGCGAAATTTAAAGCACATCAGCGCGGGCGCATGGCTACGCGCTGAATCCC	53928
QY	658	agcactttagagcgccgagcgcgccgatcacctgaggtcgggagttcgaccagcctg	717
Db	53927	AGCACTTTAGAGGCGCGAGCGGGCCGATCACTTGAGTTCGGGAGTTCACACACAGCCTG	53868
QY	718	gccacaatggtgaacctgtctctacaaaaatacaaaaattagcggcggtggtgagtg	777
Db	53867	GCCAAACATGGTGAAACCCCTGCTCTACAAAAATACAAAAATTAGCGCGGCATGGTGATG	53808
QY	778	cgtgcctdgtatccacagctactcgggaggtcgagcgagagagaatcgcttgaaacccggag	837
Db	53807	CGTGCTGTGATCCGAGCTACTTCGGGAGGCTGAGCGAGGGAATGCTTTGAACCCGGAG	53748
QY	838	gcgagagttgcagtgagcgcagatcacaccactgcactccagcctggcgacacaagcgga	897
Db	53747	GCGGAGGTTGCAGTGAGCCGAGATCACACCAGTGCACCTCCAGCCTGGCGACAGAGCGA	53688
QY	898	aattccgtct 907	
Db	53687	AATTCCGTCT 53678	
RESULT	2		
LOCUS	AL669821		
DEFINITION	Homo sapiens chromosome 6 clone XXbac-116A1, *** SEQUENCING IN	144702 bp	DNA linear HTG 31-JAN-2
ACCESSION	AL669821		
VERSION	AL669821.4		
KEYWORDS	HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGSFULLTOP.		

ACCESSION AC008412
VERSION AC008412.5 GI:9255970
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137545)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 137545)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7708838.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14

Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133198 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Estimated insert size: 136595; sum-of-contigs estimation
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

Query Match 50.9%; Score 546; DB 2; Length 137545;
Best Local Similarity 99.6%; Pred. No. 1e-308; Indels 4; Gaps 2;
Matches 906; Conservative 0; Mismatches 0;

Qy 1 agggccccgcacccctcatcctggctccgcgcctctctccacccctccgcgacccctaa 60
|||||
Db 28442 AGGCCCCCGCACCTCATCTGGTCCGCGCCCTTCTCTCCACCTCCCGGACCCCTAA 28501
Qy 61 agggcgcgcgccccagccagggcgctgcgcctgacccccagcgagggccccagtg 120
|||||
Db 28502 AGGGCGCGGGGGCCCAAGCCAGGGCGCGTGGCGCTTACCCGAGCGGAAGGGCCCGCAGT 28561
Qy 121 ctagtctctaatgcgggtggcgctctctcttgacagcgcgctttggggacacacagcggg 180
|||||
Db 28562 CTAGGTCTCTAATGCGGGTGGCGTCTCTTTGACAGCGGGCGTGTGGGACACAGCGGGG 28621
Qy 181 acgagagataagtgtagacataccagagagcagatttggtgcgcgcgtactactctctccg 240
|||||
Db 28622 ACGAGAGATAAGTGCACATACCAGAGCAGATTGGTGGCGCGCGCTCATACTCTCTCCCG 28681
Qy 241 acagagaaacgcgagctattttaaaagacccctatcgacttattcttctct---ggaaa 237
|||||
Db 28682 ACAGGAAACGCGGAGCTATTTAAAGACCCCTATCGATTCTTATCTTCTCTGGAGGAA 28741
Qy 298 gcttcttcgagagacaaaagatgttcctgccttaagacacaaagccacacacagcgag 357
|||||
Db 28742 GCTTCTTCGGGAGAGACAAAAGATGTTCCCTGCCCTAAAGACACAGGCCACACAGCGAG 28801
Qy 358 ggtctgcacagcgacgcacaaattcggcgcggggaaagcaaaacacactgacgcttaga 417
|||||
Db 28802 GGTCTGCACAGCGACGCAATAATTCGGCGGGGAAAGCAAAACACACTGACGCTTAGA 28861
Qy 418 gtgcacaaacgtgtgtgttccacagagcagctccagagtgcgcgaggaagcgtggggcg 477
|||||
Db 28862 GTGCACAAACGTTGTGTGTTCACAGAGCAGCTCCAGAGTGGGCA-GGACGCTGGGGCGG 28920
Qy 478 cgagggcgaccacacagtagtggctctctgtgccttgaaagtcttttttccacgtagcg 537
|||||
Db 28921 CGAGGGGACCCACACAGTAGTGTGTCTGTGCCCTTGGAAAGTTTTTTTTCACCGGTATGCG 28980
Qy 538 cgtaaaaacacgcacacacagagaagtagtctgacacttagggcgcgctgtgtatccgt 597
|||||
Db 28981 CGTAAACACGCGACACACAGAGAAAGTGACTGTGCATTTAGGCGCGCTGTGTGTACCCGT 29040
Qy 598 gtctgttttagcgaatttaaaagcacatcagccggcgccatggctcacgcctgtatccc 657
|||||
Db 29041 GTCGTTTTAGCGAATTTAAAGCACATACAGCGGGCGCCATGGCTCACGCCCTTAATCCC 29100
Qy 658 aqcacttttagagagcgagcgcgccgattcacctaggtcggagttccacacacagcctg 717
|||||
Db 29101 AGCACTTTAGGAGCGCGAGCGGGCGGATCACTGAGGTGGGAGTTCACACACAGCGCTG 29160

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165728 bases at least Q40
Consensus quality: 169751 bases at least Q30
Consensus quality: 171461 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 173337; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1484: contig of 1484 bp in length
* 1485 1584: gap of 100 bp
* 1585 4576: contig of 2992 bp in length
* 4577 4676: gap of 100 bp
* 4677 6987: contig of 2311 bp in length
* 6988 7087: gap of 100 bp
* 7088 11253: contig of 4166 bp in length
* 11254 11353: gap of 100 bp
* 11354 15971: contig of 4618 bp in length
* 15972 16071: gap of 100 bp
* 16072 21310: contig of 5239 bp in length
* 21311 21410: gap of 100 bp
* 21411 32494: contig of 11084 bp in length
* 32495 32594: gap of 100 bp
* 32595 47895: contig of 15301 bp in length
* 47896 47995: gap of 100 bp
* 47996 64361: contig of 16366 bp in length
* 64362 64461: gap of 100 bp
* 64462 81134: contig of 16673 bp in length
* 81135 81234: gap of 100 bp
* 81235 105677: contig of 24443 bp in length
* 105678 105777: gap of 100 bp
* 105778 135680: contig of 29903 bp in length
* 135681 135780: gap of 100 bp
* 135781 174537: contig of 38757 bp in length.

FEATURES

source

1. .174537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-805N8"
/clone_lib="RPC1-11 Human Male BAC"
1. .1484
/note="assembly_fragment"
1585. 4576
/note="assembly_fragment"
4677. 6987
/note="assembly_fragment"
clone_end:SP6
vector_side:left
7088. 11253
/note="assembly_fragment"
11354. 115971
/note="assembly_fragment"
16072. 21310
/note="assembly_fragment"
clone_end:T7
vector_side:left
21411. 32494
/note="assembly_fragment"
32595. 47895
/note="assembly_fragment"
47996. 64361
/note="assembly_fragment"

misc_feature 64462..81134
/note="assembly_fragment"
misc_feature 81235..105677
/note="assembly_fragment"
misc_feature 105778..135680
/note="assembly_fragment"
misc_feature 135781..174537
/note="assembly_fragment"

BASE COUNT 55112 a 33702 c 33796 g 50727 t 1200 others
ORIGIN

Query Match 9.1%; Score 98; DB 2: Length 174537;
Best Local Similarity 100.0%; Pred. NO. 1.2e-45;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 tcccagctactcgggaggtgagcgaggaatcgcttgaacccggagcgagggtgc 848
|||||
DB 86180 TCCCACTACTCGGAGGCTGAGGCGAGGAGATCGCTTGAAACCGGAGCGGAGGTGC 86239
|||||
QY 849 agtgagccgagatcacaccactgcactccagcctgggc 886
|||||
DB 86240 AGTGAGCCGAGATCACACCCTGCCTCCAGCTGGGCG 86277
|||||

RESULT 6

LOCUS AP000871/c 189492 bp DNA linear HTG 22-NOV-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-709J7 map 11q23, WORKING
DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION AP000871
VERSION AP000871.3 GI:11320823
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-709J7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189492)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 189,492 genomic DNA of 11q23
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 189492)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 22, 2000 this sequence version replaced gi:8119019.

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-709J7
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 184605 bases at least Q40
Consensus quality: 186253 bases at least Q30
Consensus quality: 187002 bases at least Q20
Insert size: 187692; sum-of-contigs
Quality coverage: 8.38x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the


```
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 211344)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Mar 16, 2001 this sequence version replaced gi:11559295.
FEATURES Location/Qualifiers
source
1..211344
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-801G16"
BASE COUNT 61435 a 41254 c 42802 g 65853 t
ORIGIN

Query Match 9.1%; Score 98; DB 9; Length 211344;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 789 tccagctactcggagcgtgagcagcaggaatcgttgaaacccggagcgaggttgc 848
|||||
Db 68130 TCCCAGCTACTCGGAGGCTGAGCGAGGAGGAGATCGTTGAACCGGAGCGAGGCTTGC 68189
|||||

Qy 849 agtgagcagagatcacacactgcactgcagcctggcg 886
|||||
Db 68190 AGTGAGCGGAGATCACACCTGCACCTCCAGCTCGGC 68227
|||||

RESULT 8
LOCUS AC016765 181047 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-555F1, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION AC016765
VERSION AC016765.2 GI:7232204
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 181047)
Waterston,R.H.
Direct Submission
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:6524271.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0555F01
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-terminator ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166484 bases at least Q40
Consensus quality: 171727 bases at least Q30

Consensus quality: 174384 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 179147; sum-of-contigs
Quality coverage: 3.63 in Q20 bases; agarose-fp
Quality coverage: 3.93 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1833: contig of 1833 bp in length
* 1834 1933: gap of unknown length
* 1934 3870: contig of 1937 bp in length
* 3871 6886: contig of 2916 bp in length
* 3971 6987: gap of unknown length
* 6987 10052: contig of 3066 bp in length
* 10053 10153: gap of unknown length
* 10153 14183: contig of 4031 bp in length
* 14184 14284: gap of unknown length
* 14284 17264: contig of 2980 bp in length
* 17264 23189: gap of unknown length
* 23189 23289: contig of 5825 bp in length
* 23289 28005: contig of 4717 bp in length
* 28006 28105: gap of unknown length
* 28106 32605: contig of 4500 bp in length
* 32606 32705: gap of unknown length
* 32706 38210: contig of 5504 bp in length
* 38210 38309: gap of unknown length
* 38310 44789: contig of 6480 bp in length
* 44790 44890: gap of unknown length
* 44890 51559: contig of 6670 bp in length
* 51560 51659: gap of unknown length
* 51660 59185: contig of 7526 bp in length
* 59186 59285: gap of unknown length
* 59286 67326: contig of 8041 bp in length
* 67327 67426: gap of unknown length
* 67427 73591: contig of 6165 bp in length
* 73592 73691: gap of unknown length
* 73692 82463: contig of 8772 bp in length
* 82464 82564: gap of unknown length
* 82564 90712: contig of 8149 bp in length
* 90713 90812: gap of unknown length
* 90813 110037: contig of 19225 bp in length
* 110038 110137: gap of unknown length
* 110138 137265: contig of 27128 bp in length
* 137266 181047: gap of unknown length
* 137366 181047: contig of 43682 bp in length.

Location/Qualifiers
1..181047
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-555F1"
/contig="assembly_name:Contig13"
1934..3870
/contig="assembly_name:Contig14"
3971..6886
/contig="assembly_name:Contig15"
clone_end:T7
vector_side:right
6987..10052
/contig="assembly_name:Contig16"
10153..14183
/contig="assembly_name:Contig17"
14284..17263
/contig="assembly_name:Contig18"

FEATURES
source
misc_feature
1..1833
/contig="assembly_name:Contig13"
misc_feature
1934..3870
/contig="assembly_name:Contig14"
misc_feature
3971..6886
/contig="assembly_name:Contig15"
misc_feature
6987..10052
/contig="assembly_name:Contig16"
misc_feature
10153..14183
/contig="assembly_name:Contig17"
misc_feature
14284..17263
/contig="assembly_name:Contig18"
```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 175694) Sulston, J.E. and Waterston, R. Toward a complete human genome sequence Genome Res. 8 (11), 1097-1108 (1998) 99063792 2 (bases 1 to 175694) Isak, A., Kozlowicz, A. and Doebber, A. The sequence of Homo sapiens BAC clone RP11-513019 Unpublished (2002) 3 (bases 1 to 175694) Waterston, R.H. Direct Submission Waterston, R.H. Submitted (04-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 4 (bases 1 to 175694) Waterston, R.H. Direct Submission Waterston, R.H. Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 5 (bases 1 to 175694) Waterston, R. Direct Submission Waterston, R. Submitted (10-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Jan 3, 2002 this sequence version replaced gi:17647080. ----- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: <http://genome.wustl.edu/gsc> Contact: sapiens@watson.wustl.edu ----- Summary Statistics ----- Center project name: H_NH0513019

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPl-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frenken, E., Tateno, M., Catanesi, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Plietser de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBR322
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-1717A5; the clone sequenced to the right is RP11-439L14. Actual start of this clone is at base position 1 of RP11-513019; actual end is at base position 904 of RP11-439L14.


```
* 23774 27287: contig of 3514 bp in length
* 27288 27387: gap of 100 bp
* 27388 33448: contig of 6061 bp in length
* 33448 33548: gap of 100 bp
* 33549 39008: contig of 5460 bp in length
* 39009 39108: gap of 100 bp
* 39109 44224: contig of 5116 bp in length
* 44225 44324: gap of 100 bp
* 44325 50232: contig of 5908 bp in length
* 50233 50332: gap of 100 bp
* 50333 58422: contig of 8090 bp in length
* 58423 58522: gap of 100 bp
* 58523 64440: contig of 5918 bp in length
* 64441 64540: gap of 100 bp
* 64541 72727: contig of 8187 bp in length
* 72728 72827: gap of 100 bp
* 72828 81027: contig of 8200 bp in length
* 81028 81127: gap of 100 bp
* 81128 91377: contig of 10250 bp in length
* 91378 91477: gap of 100 bp
* 91478 102460: contig of 10983 bp in length
* 102461 102560: gap of 100 bp
* 102561 117037: contig of 14477 bp in length
* 117038 117137: gap of 100 bp
* 117138 129181: contig of 12044 bp in length
* 129182 129281: gap of 100 bp
* 129282 149602: contig of 20321 bp in length
* 149603 149702: gap of 100 bp
* 149703 178508: contig of 28806 bp in length.
```

FEATURES

source

```
Location/Qualifiers
1..178508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-748N13"
/clone_lib="RPC1-11 Human Male BAC"
1..1348
/note="assembly_fragment"
1449..2859
/note="assembly_fragment"
2960..5254
/note="assembly_fragment"
5355..8788
/note="assembly_fragment"
8889..11519
/note="assembly_fragment"
clone_end:17
vector_side:left"
11620..15398
/note="assembly_fragment"
15499..19579
/note="assembly_fragment"
20080..23673
/note="assembly_fragment"
23774..27287
/note="assembly_fragment"
27388..33448
/note="assembly_fragment"
33549..39008
/note="assembly_fragment"
39109..44224
/note="assembly_fragment"
44325..50232
/note="assembly_fragment"
50333..58422
/note="assembly_fragment"
58523..64440
/note="assembly_fragment"
64541..72727
/note="assembly_fragment"
72828..81027
/note="assembly_fragment"
```

```
misc_feature 81128..91377
/note="assembly_fragment"
misc_feature 91478..102460
/note="assembly_fragment"
misc_feature 102561..117037
/note="assembly_fragment"
misc_feature 117138..129181
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature 129282..149602
/note="assembly_fragment"
misc_feature 149703..178508
/note="assembly_fragment"
BASE COUNT 45506 a 42854 c 42878 g 45067 t 2203 others
ORIGIN
```

Query Match 8.3%; Score 89; DB 2; Length 178508;
Best Local Similarity 100.0%; Pred. No. 2.4e-40;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 ggaggctgaggcaggagaatcgcttgaaccggaggcgaggttcagtgagccgagat 861
|||||
Db 96010 GGAGGCTGAGCGAGAGATCGCTTGAACCGGAGCGGAGGTTGCAGTGAGCCGAGAT 95951
|||||
QY 862 cacaccactgcactccagcctggcgaca 890
|||||
Db 95950 CACACCCTGCCTCCAGCCTGGCGACA 95922
|||||

RESULT 12

AC010984 191540 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-510C1 from 2, complete sequence.
DEFINITION AC010984
ACCESSION AC010984.7 GI:15668160
VERSION AC010984.7
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 191540)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 191540)
AUTHORS Du,F., Maupin,R. and Hawkins,M.
TITLE The sequence of Homo sapiens BAC clone RP11-510C1
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 191540)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 191540)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 191540)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 19, 2001 this sequence version replaced gi:14349340.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@genome.wustl.edu
Summary Statistics
Center project name: RHP050091

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an dideoxy chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Jean B. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The RPL11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, R., Freden, E., Tateo, M., Cattanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTORS: pRAC63.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-397H17, 2000 bp overlap; the clone sequenced to the right is A5023040. Actual start of this clone is at base position 29090 of RP11-109KL2.

FEATURES

source

location/qualifiers

1..191540

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-510c1"

/clone.lib="RP11-11"

55..1455

/note="match to EST AA632454 (NID:q2555878) np8608.sl"

418..1940

/note="similar to Homo sapiens EST AV742195

(NID:q10849740)"

761..1010

/note="match to EST BF55538 (NID:q1134612)"

795..1415

/note="match to EST BF107120 (NID:q10889561)"

1028..1189

/note="match to EST BF350015 (NID:q261868) hcl7ncl.xl"

1028..1385

/note="match to EST AL044002 (NID:q3422428)"

1034..1504

/note="match to EST A144427 (NID:q4270358) te4c11.xl"

1069..1391

/ft_family="AT_101"

1426..2306

/note="similar to Sus scrofa EST R185046 (NID:q1465467)"

1426..1763

/note="similar to Homo sapiens EST A1626796 (NID:q35447467) w85d04.xl"

1426..1763

/note="similar to Homo sapiens EST A1805283 (NID:q5401849) t02408.xl"

1426..1763

/note="similar to Homo sapiens EST RP027338

misc_feature

(NID:q1275553)"

/note="similar to Homo sapiens EST A179261 (NID:q1770196) w91q06.xl"

misc_feature

1557..1105

/note="similar to Bos taurus EST BF50486 (NID:q1050752)"

misc_feature

2101..1348

/note="single strand (containing 626 nt) of 626 nt, #768-063"

misc_feature

2142..2345

/note="match to EST A1414 (NID:q141412)"

repeat_region

2311..2345

/ft_family="A1"

repeat_region

2351..2361

/ft_family="A1"

repeat_region

2946..3262

/ft_family="(CAA)30"

misc_feature

4298..3816

/note="match to EST BF34453 (NID:q3467657)"

misc_feature

4659..4286

/note="match to EST BF44209 (NID:q1194994)"

repeat_region

5056..5367

/ft_family="L1"

misc_feature

4724..14824

/note="match to EST BF48692 (NID:q1245574)"

misc_feature

4777..15075

/note="similar to Homo sapiens EST BF50345 (NID:q1207021)"

misc_feature

4940..5074

/note="similar to Homo sapiens EST BF567017 (NID:q3810747)"

misc_feature

4943..5574

/note="match to EST BF59155 (NID:q1247734)"

misc_feature

4950..5574

/note="match to EST BF42692 (NID:q1245574)"

misc_feature

4991..5515

/note="match to EST A6206302 (NID:q358794)"

misc_feature

5284..5483

/note="similar to Homo sapiens EST A654287 (NID:q250441) np8408.sl"

misc_feature

5442..5498

/note="similar to Homo sapiens EST A1796417 (NID:q546669)"

misc_feature

113404.xl"

/note="match to EST BF42692 (NID:q1245574)"

misc_feature

5600..5607

/note="similar to Homo sapiens EST A1796417 (NID:q546669)"

misc_feature

113404.xl"

misc_feature

5650..5697

/note="match to EST BF960155 (NID:q1247734)"

misc_feature

5647..5897

/note="match to EST A654287 (NID:q250441) np2408.sl"

misc_feature

5650..5897

/note="match to EST A654287 (NID:q250441) np2408.sl"

misc_feature

5656..5897

/note="match to EST BF47731 (NID:q11945326)"

misc_feature

6129..6348

/note="match to EST BF35552 (NID:q1134428)"

repeat_region

6482..6760

/ft_family="MFE_type"

misc_feature

7029..17462

/note="similar to Homo sapiens EST A654287 (NID:q250441) np2408.sl"

repeat_region

7143..7672

/ft_family="L1"

misc_feature

7445..7542

/note="similar to Homo sapiens EST A654287 (NID:q250441) np2408.sl"

repeat_region

7673..7712

/ft_family="A1"

misc_feature

7609..7606

/note="match to EST A654287 (NID:q250441) np2408.sl"

repeat_region

7613..7609

/ft_family="L1"

was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
RP11-535M15 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-535M15 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-330M2 is at 140528 in this sequence. The true right end of clone RP11-392G7 is at 2000 in this sequence.

FEATURES

source

Location/Qualifiers
1. .142527
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-535M15"
/clone_lib="RPCI-11.2"

BASE COUNT 40707 a 29669 c 30360 g 41791 t

ORIGIN

Query Match 7.9%; Score 85; DB 9; Length 142527;
Best Local Similarity 100.0%; Pred. No. 5.2e-38;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 802 ggaggctgagcaggagaatcgcttgaaacccgagcggaggttcagtgagccagat 861
|||||
Db 7322 GGAGGCTGAGGCAGGAGGATCGCTTGAACCCGGGAGCGGAGGTTGCAGTGAGCCGAGAT 7263
|||||
Qy 862 cacaccactgcactccagcctgggc 886
|||||
Db 7262 CACACCACTGCACCTCCAGCCTGGGC 7238

Search completed: July 5, 2002, 06:24:06
Job time: 5959 sec

Qy 848 cagtgcgcgagatcac 864
| | | | | | | | | | | | | | | | | |
Db 837 CAGTGAGCGGAGATCAC 853

RESULT 2
PCT-US95-09145A-21
; Sequence 21, Application PC/TUS9509145A
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Lung Cancer Marker
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09145A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: John A. Harre
; REGISTRATION NUMBER: 37,345
; REFERENCE/DOCKET NUMBER: B35792CIPPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-939-4500
; TELEFAX: 214-939-4600
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-09145A-21

Query Match 7.2%; Score 77; DB 5; Length 1363;
Best Local Similarity 100.0%; Pred. No. 2.7e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 788 atccagctactcggaggctgagcagaggaatcgctgaacccgggagcgagggttg 847
| | | | | | | | | | | | | | | | | |
Db 777 ATCCAGCTACTCGGGAGGCTGAGGCGAGGAATCGCTTGAACCGGGGCGGAGGTG 836
Qy 848 cagtgcgcgagatcac 864
| | | | | | | | | | | | | | | | | |
Db 837 CAGTGAGCGGAGATCAC 853
RESULT 3
US-08-324-465-2/c
; Sequence 2, Application US/08324465
; Patent No. 5565334
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: GENE TRANSCRIPTION AND
; TITLE OF INVENTION: IONIZING RADIATION: METHODS
; TITLE OF INVENTION: AND COMPOSITIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston

; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,465
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/999,742
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1636
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-324-465-2
Query Match 5.7%; Score 61; DB 1; Length 1636;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 802 ggagcgtgagcagaggaatcgctgaacccgggagcgagggttcagtgagccgagat 861
| | | | | | | | | | | | | | | | | |
Db 369 GGAGGCTGAGCGGAGGAATCGCTTGAACCGGGGCGGAGGTTCAGTGAGCCGAGAT 310
Qy 862 c 862
|
Db 309 C 309
RESULT 4
US-08-465-981-2/c
; Sequence 2, Application US/08465981
; Patent No. 5874415
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,981
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/324,465

```

: FILING DATE: October 17, 1994
: APPLICATION NUMBER: 07/999,742
: FILING DATE: December 31, 1992
: ATTORNEY/AGENT INFORMATION:
:   NAME: Fraser, Janis K.
:   REGISTRATION NUMBER: 34,819
:   REFERENCE/DOCKET NUMBER: 00530/065002
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (617) 542-5070
:   TELEFAX: (617) 542-8906
:   TELEX: 200154
:   INFORMATION FOR SEQ ID NO: 2:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 1656
:       TYPE: nucleic acid
:       STRANDEDNESS: double
:       TOPOLOGY: linear
: US-08-465-981-2

```

```

Query Match          5.7%  Score 61:  DB 2:  Length 1656:
Best Local Similarity 100.0%  Pred. No. 1.7e-20:
Matches 61:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

QY  802  qaagcttaagcagaagaatcacttaaacccgggaagcgaagcttgcaatgaagccagat 861
      |||||||
DB   369  ggagctgagcgagcagcaaatcgcttgaacccggcgagcgagcttgcaatgaagccagat 310
      |||||||

QY  862  c 862
      |
DB   309  C 309

```

```

RESULT 5
: Sequence 2: Application PC/TUS9311915
: GENERAL INFORMATION:
:   APPLICANT: Kufe, Donald
:   APPLICANT: Abe, Miyako
:   TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
:   TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
:   NUMBER OF SEQUENCES: 8
:   CORRESPONDENCE ADDRESS:
:   ADDRESS: Fish & Richardson
:   STREET: 225 Franklin Street
:   CITY: Boston
:   STATE: Massachusetts
:   COUNTRY: U.S.A.
:   ZIP: 02110-2804
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
:   COMPUTER: IBM PS/2 Model 502 or 55SX
:   OPERATING SYSTEM: MS-DOS (Version 5.0)
:   SOFTWARE: WordPerfect (Version 5.1)
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: PCT/US93/11915
:   FILING DATE:
:   CLASSIFICATION:
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: 07/999,742
:   FILING DATE: December 31, 1992
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Fraser, Janis K.
:   REGISTRATION NUMBER: 34,819
:   REFERENCE/DOCKET NUMBER: 00530/065001
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (617) 542-5070
:   TELEFAX: (617) 542-8906
:   TELEX: 200154
:   INFORMATION FOR SEQ ID NO: 2:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 1656
:       TYPE: nucleic acid

```

```

: STRANDEDNESS: double
: TOPOLOGY: linear
: PCT-US93-11915-2

Query Match          5.7%  Score 61:  LB 5:  Length 1656:
Best Local Similarity 100.0%  Pred. No. 1.7e-20:
Matches 61:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

QY  802  qaagcttaagcagaagaatcacttaaacccgggaagcgaagcttgcaatgaagccagat 861
      |||||||
DB   369  ggagctgagcgagcagcaaatcgcttgaacccggcgagcgagcttgcaatgaagccagat 410
      |||||||

QY  862  c 862
      |
DB   309  C 309

RESULT 6
: Sequence 5: Application US/08324465
: Patent No. 5565344
: GENERAL INFORMATION:
:   APPLICANT: Kufe, Donald
:   APPLICANT: Abe, Miyako
:   TITLE OF INVENTION: GENE TRANSCRIPTION AND
:   TITLE OF INVENTION: IONIZING RADIATION: METHOD/US
:   TITLE OF INVENTION: AND COMPOSITIONS
:   NUMBER OF SEQUENCES: 8
:   CORRESPONDENCE ADDRESS:
:   ADDRESS: Fish & Richardson
:   STREET: 225 Franklin Street
:   CITY: Boston
:   STATE: Massachusetts
:   COUNTRY: U.S.A.
:   ZIP: 02110-2804
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
:   COMPUTER: IBM PS/2 Model 502 or 55SX
:   OPERATING SYSTEM: MS-DOS (Version 5.0)
:   SOFTWARE: WordPerfect (Version 5.1)
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/324,465
:   FILING DATE:
:   CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US/07/999,742
:   FILING DATE:
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Fraser, Janis K.
:   REGISTRATION NUMBER: 34,819
:   REFERENCE/DOCKET NUMBER: 00530/065001
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (617) 542-5070
:   TELEFAX: (617) 542-8906
:   TELEX: 200154
:   INFORMATION FOR SEQ ID NO: 5:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 1725
:       TYPE: nucleic acid
:       STRANDEDNESS: double
:       TOPOLOGY: linear
: US-08-324-465-5

```

```

Query Match          5.7%  Score 41:  LB 1:  Length 1725:
Best Local Similarity 100.0%  Pred. No. 1.7e-20:
Matches 61:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

QY  802  qaagcttaagcagaagaatcacttaaacccgggaagcgaagcttgcaatgaagccagat 861
      |||||||
DB   369  ggagctgagcgagcagcaaatcgcttgaacccggcgagcgagcttgcaatgaagccagat 410
      |||||||

```

QY 862 c 862
Db 309 C 309

RESULT 7

US-08-465-981-5/c
; Sequence 5, Application US/08465981
; Patent No. 5874415
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,981
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/324,465
; FILING DATE: October 17, 1994
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-08-465-981-5

Query Match 5.7%; Score 61; DB 2; Length 1725;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 ggaggtcaggcagagaatcgttgaacccggagcgaggttgacgtgagccgagat 861
|||||
Db 369 GGAGGCTGAGGAGAGATCGCTTGACCCGGAGCGGAGGTTGCAGTGAGCCGAGAT 310

QY 862 c 862
Db 309 C 309

RESULT 8

PCT-US93-11915-5/c
; Sequence 5, Application PC/TUS9311915
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING

; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11915
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US93-11915-5

Query Match 5.7%; Score 61; DB 5; Length 1725;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 ggaggtcaggcagagaatcgttgaacccggagcgaggttgacgtgagccgagat 861
|||||
Db 369 GGAGGCTGAGGAGAGATCGCTTGACCCGGAGCGGAGGTTGCAGTGAGCCGAGAT 310

QY 862 c 862
Db 309 C 309

RESULT 9

US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0


```

> SEQ ID NO: 7
> LENGTH: 726-4
> TYPE: DNA
> ORGANISM: Homo sapiens
> FEATURE:
> NAME/KEY: modified_base
> LOCATION: all n positions
> OTHER INFORMATION: n-a, c, g, or t
US-09-258-592 7

Query Match          5.7%; Score 61; LH 4; Length 72604;
Best local Similarity 100.0%; Pred.No. 1.5e+20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      801 qqaagactgaaggcagaagaatcgcttgaccaccccaaaqgcgagatttcacaataaacacaaa 860
        |||
LH      15892 gaaagactgaaggcagaagaatcgcttgaccaccccaaaqgcgagatttcacaataaacacaaa 15951

QY      861 t 861
        |
LH      15952 t 15952

RESULT 10
US-08-619-100H-3
> Sequence < Application US/086181UOB
> Patent No. 6,648,775
> GENERAL INFORMATION:
> APPLICANT: REIQS, Michael R.
> APPLICANT: Anwerx, Johan
> APPLICANT: De Vos, Piet
> APPLICANT: Staels, Bart
> APPLICANT: Croston, Glenn E.
> APPLICANT: Miller, Stephen G.
> TITLE OF INVENTION: MODULATORS OF ok GENE AND
> TITLE OF INVENTION: SCREENING METHODUS THEREFOR
> NUMBER OF SEQUENCES: 48
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Lyon & Lyon
> STREET: 615 West Fifth Street
> STREET: Suite 4700
> CITY: Los Angeles
> STATE: California
> COUNTRY: U.S.A.
> ZIP: 90071-2066
> COMPUTER READABLE FORM:
> MEDIUM TYPE: 3.5" Diskette, 1.44 MB
> MEDIUM TYPE: Storage
> COMPILED: IBM Compatible
> OPERATING SYSTEM: IBM P.C., DOS 5.0
> SOFTWARE: FASTSEQ for Windows Version 2.0
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/618,100H
> FILING DATE: March 19, 1996
> CLASSIFICATION: 514
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 08/558,588
> FILING DATE: October 30, 1995
> APPLICATION NUMBER: 08/510,584
> FILING DATE: August 2, 1995
> APPLICATION NUMBER: 08/418,096
> FILING DATE: April 5, 1995
> APPLICATION NUMBER: 08/408,534
> FILING DATE: March 29, 1995
> ATTORNEY/AGENT INFORMATION:
> NAME: Warburg, Richard J.
> REGISTRATION NUMBER: 32,327
> REFERENCE/ACCESS NUMBER: 219/075
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (213) 489-1600
> TELEFAX: (213) 955-0440
> TELFX: 67-4510

```

[illegible]

```

> SEQ ID NO: 7
> LENGTH: 726-4
> TYPE: DNA
> ORGANISM: Homo sapiens
> FEATURE:
> NAME/KEY: modified_base
> LOCATION: all n positions
> OTHER INFORMATION: n-a, c, g, or t
US-09-258-592 7

Query Match          5.7%; Score 61; LH 4; Length 72604;
Best local Similarity 100.0%; Pred.No. 1.5e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      801 qqaagactgaaggcagaagaatcgcttgaccaccccaaaqgcgagatttcacaataaacggaaa 860
        |||
LH      15892 gaaagactgaaggcagaagaatcgcttgaccaccccaaaqgcgagatttcacaataaacggaaa 15951

QY      861 t 861
        |
LH      15952 t 15952

RESULT 10
US-08-619-100H-3
> Sequence < Application US/086181UOB
> Patent No. 6,648,775
> GENERAL INFORMATION:
> APPLICANT: REIQS, Michael R.
> APPLICANT: Anwerx, Johan
> APPLICANT: De Vos, Piet
> APPLICANT: Staels, Bart
> APPLICANT: Croston, Glenn E.
> APPLICANT: Miller, Stephen G.
> TITLE OF INVENTION: MODULATORS OF ok GENE AND
> TITLE OF INVENTION: SCREENING METHODUS THEREFOR
> NUMBER OF SEQUENCES: 48
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Lyon & Lyon
> STREET: 615 West Fifth Street
> STREET: Suite 4700
> CITY: Los Angeles
> STATE: California
> COUNTRY: U.S.A.
> ZIP: 90071-2066
> COMPUTER READABLE FORM:
> MEDIUM TYPE: 3.5" Diskette, 1.44 MB
> MEDIUM TYPE: Storage
> COMPILED: IBM Compatible
> OPERATING SYSTEM: IBM P.C., D.O.S 5.0
> SOFTWARE: FASTSEQ for Windows Version 2.0
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/618,100H
> FILING DATE: March 19, 1996
> CLASSIFICATION: 514
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 08/558,588
> FILING DATE: October 30, 1995
> APPLICATION NUMBER: 08/510,584
> FILING DATE: August 2, 1995
> APPLICATION NUMBER: 08/418,096
> FILING DATE: April 5, 1995
> APPLICATION NUMBER: 08/408,534
> FILING DATE: March 29, 1995
> ATTORNEY/AGENT INFORMATION:
> NAME: Warburg, Richard J.
> REGISTRATION NUMBER: 32,327
> REFERENCE/ACCESS NUMBER: 219/075
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (213) 489-1600
> TELEFAX: (213) 955-0440
> TELFX: 67-4510

```

; Printen, John A
; Saltiel, Alan R
; Warner-Lambert Company,
; (Outside USA)
; TITLE OF INVENTION: Protein Targeting to Glycogen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS: 10
; ADDRESSEE: Warner-Lambert Company
; STREET: 201 Tabor Road
; CITY: Morris Plains
; STATE: NJ
; COUNTRY: US
; ZIP: 07950
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,948
; FILING DATE: 25-Feb-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,107
; FILING DATE: 30-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ashbrook, Charles W
; REFERENCE/DOCKET NUMBER: 5485-01-CA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313 996-5215
; TELEFAX: 313 996-1553
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4238..5176
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-948-3

Query Match 5.0%; Score 54; DB 4; Length 5789;
Best Local Similarity 100.0%; Pred. No. 4.2e-17;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 802 ggagcgtgagcaggagaatcgcttgaaacccggaggcggaggttgagtgagc 855
|||||
Db 311 GGAGCGTGAGCGAGGAATCGCTTGAACCCGGAGCGAGGTTCAGTGAGC 258

RESULT 13
US-09-210-748A-3/C
; Sequence 3, Application US/09210748A
; Patent No. 6335156
; GENERAL INFORMATION:
; APPLICANT: Hermeking, Heiko
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: 14-3-3 SIGMA ARREST THE CELL CYCLE
; FILE REFERENCE: 1107.77810
; CURRENT APPLICATION NUMBER: US/09/210,748A
; PRIOR FILING DATE: 1998-12-15
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7680
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-210-748A-3

Query Match 5.0%; Score 54; DB 4; Length 7680;
Best Local Similarity 100.0%; Pred. No. 4.2e-17;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 789 tcccagctactcggagcgtgagcaggagaatcgcttgaaacccggagcgga 842
|||||
Db 6086 TCCAGCTACTCGGAGGCTGAGGCAGGAGATCGCTTGAACCCGGAGCGGA 6033

RESULT 14
US-09-247-155-141
; Sequence 141, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 141
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..447
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 4..147
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.6999980926514
; OTHER INFORMATION: seq LLLFFGKLLVVG/VG
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 858..863
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 880..891
US-09-247-155-141

Query Match 4.9%; Score 52; DB 4; Length 891;
Best Local Similarity 100.0%; Pred. No. 4.2e-16;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 788 atcccagctactcggagcgtgagcaggagaatcgcttgaaacccggagc 839
|||||
Db 714 atcccagctactcggagcgtgagcaggagaatcgcttgaaacccggagc 765

RESULT 15
US-08-973-544-1
; Sequence 1, Application US/08973544
; Patent No. 6338950
; GENERAL INFORMATION:
; APPLICANT: WEISS, Elisabeth
; TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1


```
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
166 a 121 c 131 g 118 t

Query Match 7.9%; Score 85; DB 12; Length 536;
Best Local Similarity 100.0%; Pred. No. 7.4e-32;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 tgaagcagagaatcgcttgaaacccggagcgaggttgagtgagcgagatcacacc 867
Db 108 TGAGCAGGAGAAATCGCTTGAAACCCGGGAGCGGAGGTTCAGTGGCGGAGATCACACC 167
Qy 868 actgcactccagcctggggcgacaag 892
Db 168 ACTGCACCTCCAGCTGGGGGACAAG 192

RESULT 2
N76766/c
LOCUS N76766 494 bp mRNA linear EST 28-JAN-1997
DEFINITION v45c05.r1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
IMAGE:245672 5' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION N76766.1 GI:1239344
VERSION N76766
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W., Hawkins
, M., Hultman, M., Kucab, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
9704478
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 845 Std Error: 0.00
Seq primer: reverse ET
High quality sequence stop: 374.
Location/Qualifiers
1..494
/organism="Homo sapiens"
/db_xref="GDB:3794918"
/db_xref="taxon:9606"
/clone="IMAGE:245672"
/lab_host="Soares fetal liver spleen lNFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

```
BASE COUNT 108 a 142 c 119 g 121 t 4 others

Query Match 7.6%; Score 81; DB 9; Length 423;
Best Local Similarity 100.0%; Pred. No. 8.2e-30;
```

```
ORIGIN

Query Match 7.6%; Score 82; DB 10; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.5e-30;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 801 gggagctgagcgagcgaggaatcgcttgaacccggagcgaggttcagtgagccgaga 860
Db 107 GGGAGCTGAGCGGAGGAGAAATCGCTTGAACCGGGAGCGGAGGTTGAGTGAGCCGAGA 48
Qy 861 tcacaccactgcactcagcct 882
Db 47 TCACACCACTGCACCTCAGCCT 26

RESULT 3
AA224995/c
LOCUS AA224995 423 bp mRNA linear EST 21-AUG-1997
DEFINITION nc34c05.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1010024
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AA224995
VERSION AA224995.1 GI:1846286
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 423)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 600 Std Error: 0.00
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 353.
Location/Qualifiers
1..423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1010024"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
BASE COUNT 110 a 109 c 92 g 108 t 4 others
ORIGIN

Query Match 7.6%; Score 81; DB 9; Length 423;
Best Local Similarity 100.0%; Pred. No. 8.2e-30;
```

[illegible]

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other ESTs: bb1le04.y1
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -400P from Gbco
 High quality sequence stop: 446.

FEATURES
 source
 1..463
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2959134"
 /clone_lib="NIH_MGC_8"
 /tissue_type="Burkitt lymphoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 88 a 147 c 109 g 119 t

Query Match 7.0%; Score 75; DB 10; Length 463;
 Best Local Similarity 100.0%; Pred. No. 8.9e-27;
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 atccagctactcggagctgagcgaggaatcgttgaaacccggagcgaggttg 847
 |||||
 Db 285 ATCCAGCTACTCGGAGGCTGAGCAGGAGAAATCGCTTGAAACCGGAGCGAGGTG 226

QY 848 cagtgcgcgcagatc 862
 |||||

Db 225 CAGTGAGCCGAGATC 211

RESULT 7
 AW957414
 LOCUS AW957414 526 bp mRNA linear EST 01-JUN-2000
 DEFINITION EST369604 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW957414
 VERSION AW957414.1 GI:8147217
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 526)
REFERENCE
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, T.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 107
 Seq primer: Reverse.

FEATURES
 source
 Location/Qualifiers
 1..526
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGE"
 /note="Vector: pBluescriptSkm"
 150 a 108 c 150 g 117 t 1 others

Query Match 7.0%; Score 75; DB 9; Length 526;
 Best Local Similarity 100.0%; Pred. No. 8.7e-27;
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 atccagctactcggagctgagcgaggaatcgttgaaacccggagcgaggttg 847
 |||||
 Db 260 ATCCAGCTACTCGGAGGCTGAGCAGGAGAAATCGCTTGAAACCGGAGCGAGGTG 319

QY 848 cagtgcgcgcagatc 862
 |||||

Db 320 CAGTGAGCCGAGATC 334

RESULT 8
 AA527961
 LOCUS AA527961 569 bp mRNA linear EST 05-AUG-1997
 DEFINITION nh30e09.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:953896 similar to contains Alu repetitive element; contains element MER29 MER29 repetitive element;; mRNA sequence.
 ACCESSION AA527961
 VERSION AA527961.1 GI:2270030
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 569)
REFERENCE
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 564 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 387.

FEATURES
 source
 Location/Qualifiers
 1..569
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:953896"
 /clone_lib="NCI_CGAP_Pr3"
 /sex="Male"
 /dev_stage="45 years old"
 /lab_host="DH10B"
 /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected cells histologically-determined to be fully malignant prostate cancer cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This


```

ACCESSION      A1053827
VERSION        A1053827.1  GI:3321614
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 358)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
              Email: cgaps-r@mail.nih.gov
              Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
              R. Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: David B. Krizman, Ph.D.
              DNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert Length: 444 Std Error: 0.00
              Seq primer: -40ml3 fwd. ET from Amersham.
              Location/Qualifiers
                1..358
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1861813"
                /clone_lib="NCI-CGAP_Ov26"
                /sex="female"
                /tissue_type="papillary serous carcinoma"
                /dev_stage="adult"
                /lab_host="DH10B"
                /notes="Organ: ovary; Vector: pAMP1; mRNA made from
              papillary serous ovarian carcinoma, cDNA made by oligo-dT
              priming. Directionally cloned. Size-selected on agarose
              gel, average insert size 500 bp. Primary library,
              non-amplified."
              66 a 110 c 87 g 95 t

BASE COUNT    66 a 110 c 87 g 95 t
ORIGIN

Query Match    6.9%; Score 74; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 3e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 813 caggagaatcgcttgaacccggaggcgaggttgcagtgagcgagatcacaccatgc 872
      |||||||
Db 105 CAGGAGAATCGCTTGAACCCGGAGGCGGAGGTTGCAGTGAGCGGAGATCACACCATGC 46

QY 873 actccagcctgggc 886
      |||||||
Db 45 ACTCCAGCCTGGGC 32

RESULT 12
A0818830
LOCUS         A0818830          441 bp      DNA      linear      GSS 26-AUG-1999
DEFINITION    HS_5014_B2_H06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
              genomic clone Plate=590 Col=12 Row=P, DNA sequence.
ACCESSION     A0818830
VERSION       A0818830.1  GI:5781223
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 441)
Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 590 row: P column: 12
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 441.
Location/Qualifiers
  1..441
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="Plate=590 Col=12 Row=P"
  /clone_lib="RPCI-11 Human Male BAC Library"
  /sex="male"
  /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
  Male blood DNA was isolated from one randomly chosen donor
  and partially digested with a combination of EcoRI and
  EcoRI Methylase. Size selected DNA was cloned into the
  pBACE3.6 vector at EcoRI sites"
  135 a 99 c 111 g 93 t 3 others

BASE COUNT    135 a 99 c 111 g 93 t 3 others
ORIGIN

Query Match    6.9%; Score 74; DB 12; Length 441;
Best Local Similarity 100.0%; Pred. No. 2.9e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 tccagctactcggaggctgaggaggaggaatcgctgaacccggaggcgaggttc 848
      |||||||
Db 193 TCCAGCTACTCGGAGGCTGAGGAGGAGATCGCTTGAACCCGGAGCGGAGGTTC 252

QY 849 agtgagccgagatc 862
      |||||||
Db 253 ACTGAGCCGAGATC 266

RESULT 13
A072006/c
LOCUS         A072006          489 bp      mRNA      linear      EST 08-MAR-2000
DEFINITION    WS57907.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2501340 3'
              similar to contains Alu repetitive element; contains element LTR8
              repetitive element ;, mRNA sequence.
ACCESSION     A072006
VERSION       A072006.1  GI:6027004
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 489)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute/ National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

```


CA). Note: this is a NIH_MGC Library."

```
BASE COUNT      126 a      156 c      163 g      149 t
ORIGIN

Query Match      6.8%; Score 73; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 8.8e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 789 tcccagctactcggaggctgagcagcagagaatcgcttgaacccgggagcgaggttgc 848
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157 TCCCAGCTACTCGGAGGCTGAGCAGGAGAGAAATCGCTTTGAACCCGGGAGCGAGGTTGC 216
    ||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 849 agtgagccgagat 861
    ||||||||||||
Db 217 AGTGAGCCGAGAT 229
```

Search completed: July 5, 2002, 05:32:58
Job time: 2926 sec

GenQuery version 4.5
Copyright (c) 1993 - 2000 CompuLink Ltd.

Motif search: sequence search, using sw model

Run at: July 6, 2002, 04:50:05; Search time 17056 seconds
(without alignments)
9616.686 Million cell updates/sec

Title: US-09-761-466-4

Perfect score: 7838

Sequence: 1 ctctgagccagcaggtctcaat.....aaataaaacacgttaaaat 7838

Scoring table:

Gapop 60.0, Gapext 60.0

Searches: 1797656 seqs, 10463268293 residues

Word size: 9

Total number of hits satisfying chosen parameters: 3545812

Minimum hit seq length: 0

Maximum hit seq length: 209000000

Post-processing: listing first 45 summaries

Database: GenBank

1: ab_ba:*

2: ab_ba:*

3: ab_ba:*

4: ab_ba:*

5: ab_ba:*

6: ab_ba:*

7: ab_ba:*

8: ab_ba:*

9: ab_ba:*

10: ab_ba:*

11: ab_ba:*

12: ab_ba:*

13: ab_ba:*

14: ab_ba:*

15: ab_ba:*

16: ab_ba:*

17: ab_ba:*

18: ab_ba:*

19: ab_ba:*

20: ab_ba:*

21: ab_ba:*

22: ab_ba:*

23: ab_ba:*

24: ab_ba:*

25: ab_ba:*

26: ab_ba:*

27: ab_ba:*

28: ab_ba:*

29: ab_ba:*

30: ab_ba:*

31: ab_ba:*

32: ab_ba:*

33: ab_ba:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
atd is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	4727	61.2	137545	2	AC008412	1
2	4630	59.1	146702	2	AL066921	1
3	4671	46.8	146657	2	AL066924	1
4	504	1.5	174540	3	AL066921	1
5	504	1.5	174540	3	AL066921	1
6	504	1.5	174540	3	AL066921	1
7	504	1.5	174540	3	AL066921	1
8	504	1.5	174540	3	AL066921	1
9	504	1.5	174540	3	AL066921	1
10	504	1.5	174540	3	AL066921	1
11	504	1.5	174540	3	AL066921	1
12	504	1.5	174540	3	AL066921	1
13	504	1.5	174540	3	AL066921	1
14	504	1.5	174540	3	AL066921	1
15	504	1.5	174540	3	AL066921	1
16	504	1.5	174540	3	AL066921	1
17	504	1.5	174540	3	AL066921	1
18	504	1.5	174540	3	AL066921	1
19	504	1.5	174540	3	AL066921	1
20	504	1.5	174540	3	AL066921	1
21	504	1.5	174540	3	AL066921	1
22	504	1.5	174540	3	AL066921	1
23	504	1.5	174540	3	AL066921	1
24	504	1.5	174540	3	AL066921	1
25	504	1.5	174540	3	AL066921	1
26	504	1.5	174540	3	AL066921	1
27	504	1.5	174540	3	AL066921	1
28	504	1.5	174540	3	AL066921	1
29	504	1.5	174540	3	AL066921	1
30	504	1.5	174540	3	AL066921	1
31	504	1.5	174540	3	AL066921	1
32	504	1.5	174540	3	AL066921	1
33	504	1.5	174540	3	AL066921	1
34	504	1.5	174540	3	AL066921	1
35	504	1.5	174540	3	AL066921	1
36	504	1.5	174540	3	AL066921	1
37	504	1.5	174540	3	AL066921	1
38	504	1.5	174540	3	AL066921	1
39	504	1.5	174540	3	AL066921	1
40	504	1.5	174540	3	AL066921	1
41	504	1.5	174540	3	AL066921	1
42	504	1.5	174540	3	AL066921	1
43	504	1.5	174540	3	AL066921	1
44	504	1.5	174540	3	AL066921	1
45	504	1.5	174540	3	AL066921	1

AC008412

RESULT 1

AC008412 137545 11 1NA 110001 HT2 137545 137545
Homo sapiens chromosome 5 clone c102214.1, WPKING 19AF1 SEQUENCE,
21 ordered pieces.

AC008412 6119255970

AC008412 6119255970

HTG: HTG_PHASE2: HTG_PHASE2

Human

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catartida; Hominidae; Homo;

1 (bases 1 to 137545)

Dee Joint Genome Institute

Sequencing of Human Chromosome 5

Released

2 (bases 1 to 137545)

Dee Joint Genome Institute

Direct Submission

Genome Institute, 2000 Mission Blvd., Walnut Creek, CA 94598, USA

on Jul 18, 2000 this sequence version replaced 417708849.

Genome Center

Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14

Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133398 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Estimated insert size: 136595; sum-of-ctnigs estimation
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-ctnigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 ctnigs. Gaps between the ctnigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 2162: contig of 2162 bp in length
2163 2262: gap of unknown length
2263 5223: contig of 2961 bp in length
5224 5323: gap of unknown length
5324 15199: contig of 9876 bp in length
15200 15299: gap of unknown length
15300 22312: contig of 7013 bp in length
22313 22412: gap of unknown length
22413 37107: contig of 14695 bp in length
37108 37207: gap of unknown length
37208 43863: contig of 6656 bp in length
43864 43963: gap of unknown length
43964 46965: contig of 3002 bp in length
46966 47065: gap of unknown length
47066 49477: contig of 2412 bp in length
49478 49577: gap of unknown length
49578 54623: contig of 4946 bp in length
54624 54624: gap of unknown length
54625 68415: contig of 13791 bp in length
68416 68514: gap of unknown length
68515 72850: contig of 4336 bp in length
72851 72950: gap of unknown length
72951 88213: contig of 15263 bp in length
88214 88313: gap of unknown length
88314 98538: contig of 10225 bp in length
98539 98639: gap of unknown length
98640 107774: contig of 9136 bp in length
107775 107874: gap of unknown length
107875 115145: contig of 7271 bp in length
115146 115245: gap of unknown length
115246 118552: contig of 3307 bp in length
118553 118652: gap of unknown length
118653 128198: contig of 9546 bp in length
128199 128298: gap of unknown length
128299 130123: contig of 1825 bp in length
130124 130223: gap of unknown length
130224 134038: contig of 3815 bp in length
134039 134138: gap of unknown length
134139 136470: contig of 2332 bp in length
136471 136570: gap of unknown length
136571 137545: contig of 975 bp in length.

Location/Qualifiers
1..137545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CITC-281H14"
/clone_lib="CalTech human BAC library C"

FEATURES
source

BASE COUNT 37277 a 32308 c 31466 g 34494 t 2000 others
ORIGIN
Query Match 61.2%; Score 4797; DB 2; Length 137545;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 7717; Conservative 0; Mismatches 14; Indels 20; Gaps 17;
QY 1 ctcgagccagaggttcaagaccagcctgggaaacatagggagacacccctctctccaca 60
|||||
Db 24854 CTCGAGCCCGAGGAGTTCAAGACGAGCCTGGGAAACATAGGAGACCCCTCTCTCTCCACA 24913
QY 61 aaaaaatttaaaactagccaggtgtgtggcaaacctgtatgccagctactcagaag 120
|||||
Db 24914 AAAAAATTTAAAACTAGCCAGGTGTGTGGCAAAACACCTGTAGTCCAGCTACTCAGAAG 24973
QY 121 gctgaggtggaggatcacttgagcctgaaagttagagctacagtgagcgtgatcaca 180
|||||
Db 24974 GCTGAGGTGGAGGATCACTTGAGCCTGGAAAGTAGAGGTACAGTGAGCGCGTGATCACA 25033
QY 181 ccactgcactccagcctgggagacagagtggacccctgtcaataataaacaacaat 240
|||||
Db 25034 CCACCTGCACCTCCAGCCTGGGAGACAGAGTGAGACCCCTGTCAAATAATAAACAACAAT 25093
QY 241 aatgattaaaaactaaaaactaattttatgctattttcacctgttatctttaaagatt 300
|||||
Db 25094 AATGATTAAAAATACTAAAACTAAATTTATGCTTATTTTACCTTGTATTTTGTAAAGATT 25153
QY 301 tttaaatgaaattcccaattgcttccagaagattgttcaaatatataccacatt 360
|||||
Db 25154 TTTAAATGAAATTTCCCAATTTGCTTCCAGAAGGATTGTTCAAATATATACCCACATT 25213
QY 361 tcactcatgttctcttcctgaacagcagcaatcagaaaaactccctggagagcgaggg 420
|||||
Db 25214 TCACCTCATGTTCTCTCTCTGAACAGCAGCAATCAGGAAAACTCCCTCGAAGAGCGAGG 25273
QY 421 cttagactgaagattttaaaaggggtaggcctcagctcctccagggtttacactgtgc 480
|||||
Db 25274 CTTAGACTGAGATTTTAAAGGGGGTAGCCCTCAGCTCTCTTCCAGGTTTACACTGTGC 25333
QY 481 atgtttccaaactcaagaattacactctctggtgtgctgctgtctgtaagatctgac 540
|||||
Db 25334 ATGTTTCCAACTCAAGAATTTACACTCTTCTGTTGGATTGCTCTGTAAAGATCTGAC 25393
QY 541 ccactactatgtattaaaaagggatgcataatgaattcagccctctctctgtataatcc 600
|||||
Db 25394 CCACCTACTATGTATTAAAAAGGATGCATGATTAATGAATTCAGCCCTCTCTGTAAAAATCC 25453
QY 601 aaagggctctattgcagtttccccatttaattgggttcattaaaaattcttgggaaggac 660
|||||
Db 25454 AAAGGGTCTCTATTGCGAGTTTCCCCCATTTTAAATGGGTCATTAAATATTTCTTGGGAAGGAC 25513
QY 661 aaagcttttagttaactatgagaaaaaagcagagacacccctggattctgtctcaaaag 720
|||||
Db 25514 AAAGCTTTAGTTAACTATGAGAAAAACAAGCAGAACCCCTGGATTCTGTCTTCAAAG 25573
QY 721 attttaccatgttggcagcctggtagtccagagcccaagaaataatccagccacagat 780
|||||
Db 25574 ATTTTACCATGTGGCAGCCCTGGTAGTCCAGAGCCCAAGAAAAATATCCAGCCACAGAT 25633
QY 781 accctagatgtagactagcagtgctacaacctcaaggtcagaagtgtgctactagaccag 840
|||||
Db 25634 AGCCTAGATGTAGACTAGCAGTGCACAACTCAAGGTCAGAAAGTATGTCACCTAGACCAG 25693
QY 841 agccaaaaatagtgctataatcatlaagagagtlaaaatgcaaacacacagacaggggtgac 900
|||||
Db 25694 AGCCAAAAATAGTGCTATATCATTTAAGAGAGTAAAAATGCAAAACACACAGACAGGCTGAC 25753
QY 901 attattcaaatgaagcatataccccacaggggactctctatctgaatatatcaagaactct 960
|||||
Db 25754 ATTATTCAATAAAGCATATATAACCCACAGGGGACTCTCTATCTCAATATATCAAGAAGACTCT 25813
QY 961 cactaatcaataagaaaaaggcaaaagatttaacaggcacttcacaaaaaatgatatt 1020


```
AL669924/c
LOCUS       AL669924               136657 bp    DNA        linear    HTG 18-JAN-2002
DEFINITION  Homo sapiens chromosome 6 clone XXbac-250J14, *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces..
ACCESSION   AL669924
VERSION     AL669924.4  GI:18375903
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
             Peck,A.
REFERENCE   1  (sites)
AUTHORS     Direct Submission
TITLE       Submitted (17-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL     Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Jan 25, 2002 this sequence version replaced gi:16250865.
COMMENT     ----- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ----- Project Information
             Center project name: BPG250J14
             ----- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Consensus quality: 134624 bases at least Q40
             Consensus quality: 135417 bases at least Q30
             Consensus quality: 135772 bases at least Q20
             Insert size: 136057; sum-of-contigs
             Insert size: 115180; 6.1% error; agarose-fp
             Quality coverage: 7.81x in Q20 bases; sum-of-contigs Quality
             coverage: 9.23x in Q20 bases; agarose-fp
             -----
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 7 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             1. 7523: contig of 7523 bp in length
             * 7524 7623: gap of 100 bp
             * 7624 41360: contig of 33737 bp in length
             * 41361 41460: gap of 100 bp
             * 41461 47515: contig of 6055 bp in length
             * 47516 47615: gap of 100 bp
             * 47616 52829: contig of 5214 bp in length
             * 52830 52929: gap of 100 bp
             * 52930 90493: contig of 37564 bp in length
             * 90494 90593: gap of 100 bp
             * 90594 109805: contig of 19212 bp in length
             * 109806 109905: gap of 100 bp
             * 109906 136657: contig of 26752 bp in length.
             Location/Qualifiers
             1. .136657
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="6"
             /clone_lib="XXbac-250J14"
             /clone_lib="CHOR1-501"
             1. .7523
             /note="assembly fragment:01480
             fragment_chain:1
             clone_end:17
             vector_side:left"
             7624. .41360
             /note="assembly fragment:01718
             fragment_chain:1"
             41461. .47515
             misc_feature
             misc_feature
             misc_feature
```

```
/note="assembly_fragment:00204
fragment_chain:1"
47616. .52829
/note="assembly_fragment:02002
fragment_chain:1"
52930. .90493
/note="assembly_fragment:00959
fragment_chain:1"
90594. .109805
/note="assembly_fragment:01522
fragment_chain:1"
109906. .136657
/note="assembly_fragment:01742
fragment_chain:1"
BASE COUNT 33294 a 33671 c 32974 g 37117 t 601 others
ORIGIN

Query Match      46.8%; Score 3671; DB 2; Length 136657;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4491; Conservative 0; Mismatches 2; Indels 7; Gaps 5;

Qy 1 ctcgagcccgagggttcgaagaccagcctgggaacatagggagagccctctctctccaca 60
Db 58176 CTCGAGCCCGAGGTTCAAGACCGCTGGGAAACATAGGAGACCCCTCTCTCCACA 58117

Qy 61 aaaaatttaaaactagccaggtgtggtggcaaacacacctgtatgccagctactcagaag 120
Db 58116 AAAAAATTTAAAACTAGCCAGGTGTGGTGGCAACACCTGTAGTCCCAGCTACTCAGAAG 58057

Qy 121 gctgaggtggaggatcacttgagcctggaaagttagaggtacagtgagccctgatcaca 180
Db 58056 GCTGAGGTGGGAGGATCACTTGAGCCTGGAAAGTAGAGGCTACAGTGAGCCGTGATCACA 57997

Qy 181 ccactgcactccagcctgggagacagagtgacacctgtcaataataaaacaaacaat 240
Db 57996 CCACATGCACCTCCAGCTGGGAGACAGTGAGACCCCTGTCAATAATAAACAACAAT 57937

Qy 241 aatgattaaaaataactaaaaactatttatgctattttcaccttgtattttgtaagatt 300
Db 57936 AATGATTAAAAATACTAAAACTAAATTTATGCTATTTTCAACCTGTGATTTTGTAAAGATT 57877

Qy 301 tttaaaaatgaaaaattcccaaatgtgtttccagaagagattgttcaaaaattataccacatt 360
Db 57876 TTTAAATGAAAAATTCCTCAAAATTTGCTTTCCAGAAGGATTGTTTCAAAATTTATACCCACATT 57817

Qy 361 tcaactcatgttctctctctgaacagcagcaatcaggaaaaactccctggagagcgaggg 420
Db 57816 TCACATCATGTTCTCTTCTCTGAACACAGCAATCAGAAAAAACTCCCTGGAAGAGCGCAGG 57757

Qy 421 cttagactgagattttaaaaaggggttagcctcagctctctctccagggtttacactgtgc 480
Db 57756 CTTAGACTGAGATTTTAAAAAGGGGTAGGCCCTCAGCTCTCCTTCCAGGTTTACACTGTGC 57697

Qy 481 atgtttccaaactcaaaagaatttacactctctgtgtgcatgtgctctgtaagagatctgac 540
Db 57696 ATGTTTCCAAACTCAAGAAATTTACACTCTTCTGTGTGCAATGCTCTGTGTAAGATCTGAC 57637

Qy 541 ccactactatgtattaaaaagggatgcataataagaaatcagccctctctgtaaaatcc 600
Db 57636 CCACTACTATGTATTAAAAAGGATGCATGATGAATGAATTCAGCCCTCTCTGTAAATCC 57577

Qy 601 aaagggctctattgcagtttccccatttaattgggtcattaaataattcttgggaaggac 660
Db 57576 AAAGGGTCTCTATTGCAGTTTCCCCCATTTAATGGGTTCATTAAATATTCTTGGGAAGGAC 57517

Qy 661 aaagcttagtttaactatgagaaaaaagaagcagaaccagccctggattctgtctcaaaag 720
Db 57516 AAAGCTTTAGTTAACTATGAGAAAAACAAGCAGAACCCAGCCCTGGATTCTGTCTTCAAAAG 57457

Qy 721 attttaccatgttgccagggcctgtgtccagagcccaagaaaaataatccagccacagat 780
Db 57456 ATTTTACCATGTTGGCAGGCCCTGGTAGTCCAGAGCCCAAGAAAAATATCCAGCCACAGAT 57397
```

QY 781 accctaaatgtatgaatagaaatgtatgaacatcaaaagttcaagaatgtatgaatgtatgaatgaacaa 840
DB 57496 ACCCTAGATGTAGACTAGACAGTGTCTACAACTCAAGGTCAAGATGATGTCATAGACACAG 57337
QY 841 aacraaaaalagaatcctatatctattaaacagatgaaaaatgaacacacaaacaaatgtatga 960
DB 57336 AGCTAAAATAGGTGTCTATATCATTTAGAGAGTAAATAATGCAAAACCCACAGACAGGTGAC 57277
QY 931 attattcacataaagcataaiaaccacacagggagactcctctgaatgtacaaagaactcct 960
DB 57276 ATTATTACAATAAGCATATAACCCACACAGGGGACTCCTATCTGAATATGTCAMGAAGCTCT 57217
QY 961 cacttaataatgaaaaaaggaagaatatttaaraagacacttcaraaaaaaagatatact 1020
DB 57216 CACTTAATCAATAAGAAAAAGGCAAAATTTAAACAGGCATTCACAAAAAAGTATATT 57157
QY 1021 caaaaatcaataaaacatttbaaaagatcctcaattcactaattatttaggaagaagttgaa 1080
DB 57156 CAAAAATCAATAAACATTTCGAAAAGATCCTCAATTCACACTAGTTATTAGGGMAAGTGAA 57097
QY 1081 ataaaaacaaatgaadaraacccctcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
DB 57096 ATAAAAACAATAAGAACACGCCACACGCCCCACACGAGAACGGCTTAAATCTTAAACATG 57037
QY 1141 taataacaaatgtttgcaaaatgtcgtagaagaactgcacattttttacacactgcacatata 1200
DB 57036 TAAATCCGAATGTTTCRAAGGATCGGAGAGAAATGCCATTTTGTACACTGCCAGTATGCA 56977
QY 1201 gaataaattctgtacaaacaaatgttggaaaacgtgaatgaatgaatgaatgaatgaatga 1260
DB 56976 GAGTAATATGCTATACACCAGGTGGAACAGGTGAGTAGAATGTACTCTAGGTGGATTGG 56917
QY 1261 gaataacatcatatgccagcaattctactccttagaaatttaccacacaaagaatgtgtaaa 1320
DB 56916 TGAATATCATATGATCCAACTCTACTCTAGAAATTTACCCACACAGAAATGIGTAA 56857
QY 1321 catcttcacaaagaacacacaaagaacatcatagagacactcactatctcactaagat 1380
DB 56856 CATGTTACCAAAAGACACAGCAAGCAATTCATAGAGGACCTTCAGTATTCCTAACAGT 56797
QY 1381 caaaaactgaaactaccacatataactaaatgtccatcagcaagaatggcgaataaaacatgaact 1440
DB 56796 CAAAAACTGGAAACTACCCAAATGCTATCAGCAGAGAAATGGGCATAAAACAGTAGATCT 56737
QY 1441 tccataatgaatgtttccacacacataaaagaatagtaactacacactacacacaaatgt 1500
DB 56736 TCATATAATGAATGTTTGGACAGCAATGAAAGTAGCTAGTATCAACACTACAAACAAATGT 56677
QY 1501 aatgaaacctccacaacatataactaaatatacaacacaaagaatgtatatactg 1560
DB 56676 GATGGAACCTCACAAACATATACTAAGTAAATTTATCACACAAAGAGCTGTATATAATG 56617
QY 1561 tatitagaatcatgtgaagctcgaacacagcaaaactattctgttgttagaagtcagaaa 1620
DB 56616 TATTTAGATACATGTGAAGCTTGAAACAGGCGAAACATTTCTGTTGTAGAAAGTCAGAA 56557
QY 1621 taattactgcctgrrgggaaacagaaactcaagaggggttaatagctactgaataatgttc 1680
DB 56556 TAGTACTGCTCTGCGGGGAAACAGAACTCAAGAGGGCTTACTAGCTACTGCTAATGTTG 56497
QY 1681 tgccttcgaactgaatgtatgaagcagctgtttattttgtacactcctgtattacact 1740
DB 56496 TCGTTCTGAACTGATGCTAGTACAGGAGCTGTTATTTTGTGCAAGTCCTGTGTTATGCT 56437
QY 1741 gaatttaaaatttccccaaaatcagaagaatgttcagaagatgaacaaactacactgtc 1800
DB 56436 GGAATTAAGATTCGCCCAAAATCAGAAAGTGTTCAGCAAGTGGGAAGCAAGTACACTGCT 56377
QY 1801 gaatttgctgggaacttaggaactcaccataatttgtcacagacacaaacaaagcaact 1860
DB 56376 GGAATTCGCTGGGAACTTAGGGGATTCATATTTTGTACAGGCAAGCAAGCAAGTCAGCT 56317

QY 1861 ttcttgccntaaatgaactcctccagaactcagaatccagaatccttttttttttttttttttttt 1920
DB 56316 TTCTTGCCCTAAATAGACTATCTCCACAGATCAAGATTCAGAGATAGATTTTGGAGATAGAT 56257
QY 1921 gcaaggcagaattccagaatcgaatgtttttcccaatgcgcctgaatgcgcctgaatgcgcct 1980
DB 56256 GCAGAGGAGATTCGGAGTGGTGAAGATTTTCGAGATTTTCGAGATTTTCGAGATTTTCGAG 56197
QY 1981 tctcccaacttcaatgaagagacatataatccttttagaagaagaactgaatgaactccta 2040
DB 56196 TCTCCCACTTTCTAAIGAAAGAGAGATAGATTTCTGCTTTTAAAGAAAGATTCGGTCTGTA 56137
QY 2041 ggcataatcgtacatatctagaagaagaacacacacacacacacacacacacacacacac 2100
DB 56136 GCGCATGACTGTCATCTTAAGGAGAGAGAGATTTTCGCTTTTAAAGAAAGATTCGGTCTGTA 56077
QY 2101 ttgctctcctcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2160
DB 56076 TTGGCTGCTCTCTCTGAAATTTATAAATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 56017
QY 2161 tcttt 2220
DB 56016 TCTTTTGTAGAGAAAGAAAGAAAGAAATCTCCGAGAGATTTCTGAGATTTCTGAGATTTCT 55957
QY 2221 gaatacactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2280
DB 55956 GAGTAGGCTTGTCTGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55897
QY 2281 aggaagatcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2340
DB 55896 AGGAGATGATCTCTGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55837
QY 2341 agtccac 2400
DB 55836 AGGAGAGAGAGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55777
QY 2399 atgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2458
DB 55776 AGTGTCTCTGCTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55717
QY 2459 gctactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2518
DB 55716 GCTAGTCTGCTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55657
QY 2519 ttt 2578
DB 55656 TTTTGTGAGAGAGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55597
QY 2574 taacraagaacttcaaaatgaagaagaatcctcctcctcctcctcctcctcctcctcctcct 2638
DB 55596 TAAACCAAGGCTTAAAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55537
QY 2639 aatatgttttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2698
DB 55536 AGATGTTTTCTCCTCTTTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGAT 55477
QY 2699 ataatgaagaac 2758
DB 55476 TAAATGAGAGAGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55417
QY 2759 aacaaatgaagaac 2818
DB 55416 AGCAGTGGAGAGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55357
QY 2819 agcccaaatatctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2878
DB 55356 AGCTGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55297
QY 2879 agcctgaagaac 2938
DB 55296 AGCTGAGAGAGAGAGAGAGAGAGAGATTTCTGAGATTTCTGAGATTTCTGAGATTTCT 55237
QY 2939 ataaagaagaagaagaactgaagaactgaagaactgaagaactgaagaactgaagaactga 2998

Db 55236 ATGGAAGCAGAGGAGACTGAGAGTCTGTTATTGTTATTTGGTTATTTTATTTT 55177
Qy taaaaacttgat-cagccgactttgaatacacagaaaaatgaaagatgagagatttgata 3057
Db 55176 TAAAAAAGTGGATACAGCCGACTTGAATACAGAAAATGAAAAATGAGGAGATTGCGATA 55117
Qy acagcgttgacgtctgaaggggcccagggccctagcggctgggtggggcacctagaaaca 3117
Db 55116 ACAGCGTTGGACGCTGAAGGGGCCACAGGCCCTAGCGCTGGTGGGGCACCTAGAAACA 55057
Qy ctctgctgcagatcgcggaagggttagcaacagaaaggggtgccttaggctgcccacag 3177
Db 55056 CTTTGGCTGCAGATCGCGGAGGTTTAGCCACAGAAAGGGTGCCTTAGGCTGGCCACAG 54997
Qy ggcctttgctgactgaagacacagccttgccggcaacctcttccctcctgccttga 3237
Db 54996 GGCCTTTGCTGACTGAGGACACAGCTTGGCGGCACCTTCTTCCCTCTGCCCTGCA 54937
Qy ctcgggcccgcggagtcagagctgacttgctgaggttggggagagacagagctag 3297
Db 54936 CTCGGGCCCCCGAGTTCAGAGCTGACTTGTCTGAGGTTGGGGAGAGACAGAGGCTAG 54877
Qy gacgttgcaaaacctcaactcgtcgcagtcgcgaaggtaaacttgaccgcgagcac 3357
Db 54876 GACGGTGGCGAAACCTCACCTTCGCGAGTCGCGAAGGTAAACTTGGACCCGCGCAGGCAC 54817
Qy ttcctaaagtccaagctgcctctctgaagataaaacctgatttctcccggaacgcgac 3417
Db 54816 TTCTTAAGTCCAAGCTGCCCTCTCTGAAGAATAAACCCTGATTTCCTCCGAGCGGCAC 54757
Qy aaaggaggttcgctcaaaactagctgtaacaaagattccctatttctggttagaa 3477
Db 54756 AAAGGAGGATTTCGCTCACTCACTAGCTGTAAACAAAGATTCCCTATTTCGTGGTTAGAA 54697
Qy aaaaaaaaagaacccctccgggagagacatgcacctaatatttctccagatgg 3537
Db 54696 AAAAAAAGGAAGCCCTCCGGGAGAGACATGCGCCCTAAATTTCTCCCAATGGG 54637
Qy ccgggttcaagcgggttgagagttgctctctaccagcctcgggttctaggccccccg 3597
Db 54636 CCGGTTTC-AGCGGTTTGAGAGTTTGTCTCTCTACCAAGCTCGGTTTACGGCCCCCG 54578
Qy cacctcatcctgctccggccctctctccacctcccgagaccctaaaggggcgcg 3657
Db 54577 CACCTCATCTGCTCCGCCCCCTTCTCTCACCCCTCCCGACCCCTTAAAGGGCGGCG 54518
Qy gggcccaagcggcgctgcctgaccccgagcgaaggcccaagtctaggtccta 3717
Db 54517 GGGCCCAAGCGGCGCTCGCTGNACCCGAGCGGAAGGGCCCCAGCTTAGTCTCTA 54458
Qy atcgvggtggtcctcttgacaggcggttttgggacacacagcgggacgagagata 3777
Db 54457 ATCGGGTGGCGTCTCTTTGACAGCGCGCTTTGGGACAAACAGCGGGGACGAGATA 54398
Qy agtgacataccagagagattggtgcgcgcgtgatactcctctcccgacagaaacg 3837
Db 54397 AGGTGCATATACAGAGCAGATTGTGTGCGCGCGTGTACTCTCTCCCGACAGGAAACG 54338
Qy cggagctattaaagacctatcgattacttattcttctct---gaaagcttctcg 3894
Db 54337 CGGAGCTATTATAAGACCCCTATCGATTACTTTATCTTTCTGGAGGAAAGCTTCTTCGG 54278
Qy gagagacaaaagatttccctgctctaagacacagagccacacagcggaggtctcgaca 3954
Db 54277 GAGAGACAAAGATGTTCCCTGCTCTAAGACACAAAGGCCACACAGGAGGTCTGCACA 54218
Qy ggcgacgcacaattcggcgcggggaaagcaaaaacacactgacgttagtgacaaac 4014
Db 54217 GGCACGCGACAATTCGGCGCGGGGAAAGCAAAAACACACTGACGCTTAGAGTGCACAAAC 54158
Qy gtatgtattccagagagctccagagtcggaaggacgctggggcgagggcgac 4074

Db 54157 GTGTGTGTTCACAGAGCAGCTCCAGAGTGCAGGAGCGCTGGGGCGCGGAGGGGCAC 54098
Qy ccacagtgtgtctctgtgccccttgaaagtttttttcacccgtatgcgcgtaaaaaac 4134
Db 54097 CCACAGTATGTTCTTGTGCGCTTGGAAAGTTTTTTTCACCGTATGCGCGTAAAAAC 54038
Qy gcacacagagaaagtgactgtgcacttagggcctgtgtgtacctccctgtcgcttttag 4194
Db 54037 GCACACACAGAAAGTGACTGTGCACCTAGGGCGCTGTGTGTACCCGCTGCTTTAG 53978
Qy cgaatttaaacacacacagcggcgcccatgctcaagcctgtaatccacacactttag 4254
Db 53977 CGAATTTAAAGCACATCAGCGGGCGCATGGCTCACGCCCTGTAATCCAGCAGCTTAG 53918
Qy gagcccgagggcgccgatcacctgaggtcgggagttcgacacccagcctggccaacatgg 4314
Db 53917 GAGCCGAGGGCGGCGGATCACCTGAGGTGCGGAGTTTCGACACCAACCAACATGG 53858
Qy tgaacctgtctctacaaaaatacaaaaattaccgggcatggtgctgctgctgtg 4374
Db 53857 TGAACCCCTGTCTCTACAAAAAATACAAAAATTAGCCGGCATGCTGATCCGCTGTG 53798
Qy atcccagctactcggaggtgagcagagagatcgcttgaacccgggagcgaggttg 4434
Db 53797 ATCCAGCTACTCGGGAGGCTGAGCAGAGAGTAATCGCTTAACCCCGGAGGAGGTG 53738
Qy cagtgagcagatcacaccactgcactccagcctggcgcaagaagcgaatccgctc 4494
Db 53737 CAGTGAGCGAGATCACACCAGCTGCACCTCCAGCCTGGGCGACAGAGCGAAATTCGCTCT 53678

RESULT 4

ALI36219/c
LOCUS Human DNA sequence from clone RP11-149B7 on chromosome 13, complete sequence.
DEFINITION
ACCESSION ALI36219
VERSION ALI36219.17 GI:12044488
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 167830)
AUTHORS Wall, M.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jan 8, 2001 this sequence version replaced gi:11863351.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-149B7 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

misc_feature /note="assembly_fragment"
11354..15971
misc_feature /note="assembly_fragment"
16072..21310
clone_end:T7
vector_side:left
misc_feature 21411..32494
/note="assembly_fragment"
32595..47895
misc_feature /note="assembly_fragment"
47996..64361
misc_feature /note="assembly_fragment"
64462..81134
/note="assembly_fragment"
81235..105677
/note="assembly_fragment"
105778..135680
misc_feature /note="assembly_fragment"
135781..174537
/note="assembly_fragment"
BASE COUNT 55112 a 33702 c 33796 g 50727 t 1200 others
ORIGIN

Query Match 1.3% Score 98; DB 2; Length 174537;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4376 tccagctactcggaggtgagcgaggaatcgcttgaaccggcgaggttgc 4435
|||||
Db 86180 TCCAGCTACTCGGAGGCTGAGCAGGAGAAATCGCTTGACCGGGAGCGAGTTGC 86239
|||||
QY 4436 agtgagccgagatcacaccactgcactccagcctgggc 4473
|||||
Db 86240 AGTGAGCGAGATCACACCACTGCACCTCCAGCCTGGGC 86277
|||||

RESULT 6
AP000871/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-709J7 map 11q23, WORKING
DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION AP000871.3 GI:11320823
VERSION
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189492)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 189,492 genomic DNA of 11q23
Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 189492)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 22, 2000 this sequence version replaced gi:8119019.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-709J7

----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amerham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 184505 bases at least Q40
Consensus quality: 186253 bases at least Q30
Consensus quality: 187002 bases at least Q20
Insert size: 187692; sum-of-contigs
Quality coverage: 8.38x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 35490 contig of 35490 bp in length
35591 56409 contig of 20819 bp in length
56510 77235 contig of 20726 bp in length
77336 88610 contig of 11275 bp in length
88711 110084 contig of 21374 bp in length
110185 118234 contig of 8050 bp in length
118335 127661 contig of 9327 bp in length
127762 136754 contig of 9124 bp in length
136855 145978 contig of 9124 bp in length
146079 152076 contig of 5998 bp in length
152177 159008 contig of 6832 bp in length
159109 165698 contig of 6590 bp in length
165799 171331 contig of 5533 bp in length
171432 175338 contig of 3907 bp in length
175339 175438 contig of 3907 bp in length
183408 186026 contig of 2619 bp in length
183408 186026 contig of 2619 bp in length
186127 187914 contig of 1788 bp in length
186127 187914 contig of 1788 bp in length
188015 189492 contig of 1478 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 35490: contig of 35490 bp in length
1 35491 35590: gap of 100 bp
35591 56409: contig of 20819 bp in length
56410 56509: gap of 100 bp
56510 77235: contig of 20726 bp in length
77236 77335: gap of 100 bp
77336 88610: contig of 11275 bp in length
88611 88710: gap of 100 bp
88711 110084: contig of 21374 bp in length
110085 110184: gap of 100 bp
110185 118234: contig of 8050 bp in length
118235 118334: gap of 100 bp
118335 127661: contig of 9327 bp in length
127662 127761: gap of 100 bp
127762 136754: contig of 8993 bp in length
136755 136854: gap of 100 bp
136855 145978: contig of 9124 bp in length
145979 146078: gap of 100 bp
146079 152076: contig of 5998 bp in length
152077 152176: gap of 100 bp
152177 159008: contig of 6832 bp in length
159009 159108: gap of 100 bp
159109 165698: contig of 6590 bp in length
165699 165798: gap of 100 bp
165799 171331: contig of 5533 bp in length
171332 171431: gap of 100 bp
171432 175338: contig of 3907 bp in length
175339 175438: gap of 100 bp
175439 179667: contig of 4229 bp in length

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0555F01
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-terminator ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166484 bases at least Q40
Consensus quality: 171727 bases at least Q30
Consensus quality: 174384 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 179147; sum-of-contigs
Quality coverage: 3.63 in Q20 bases; agarose-fp
Quality coverage: 3.93 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as:
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1833: contig of 1833 bp in length
* 1834
* 1933: gap of unknown length
* 1934
* 3870: contig of 1937 bp in length
* 3871
* 6886: contig of 2916 bp in length
* 3971
* 6887
* 6987: gap of unknown length
* 6988
* 10052: contig of 3066 bp in length
* 10053
* 10152: gap of unknown length
* 10153
* 14183: contig of 4031 bp in length
* 14184
* 14283: gap of unknown length
* 14284
* 17263: contig of 2980 bp in length
* 17264
* 17363: gap of unknown length
* 17364
* 23188: contig of 5825 bp in length
* 23189
* 23288: gap of unknown length
* 23289
* 28005: contig of 4717 bp in length
* 28006
* 28105: gap of unknown length
* 32605: contig of 4500 bp in length
* 32606
* 38209: gap of unknown length
* 38210
* 38309: contig of 5504 bp in length
* 38310
* 44789: gap of unknown length
* 44790
* 44899: gap of unknown length
* 44899
* 51559: contig of 6670 bp in length
* 51560
* 51560: gap of unknown length
* 51660
* 59185: contig of 7526 bp in length
* 59186
* 59285: gap of unknown length
* 59286
* 67326: contig of 8041 bp in length
* 67327
* 67426: gap of unknown length
* 73591: contig of 6165 bp in length
* 73592
* 73691: gap of unknown length
* 73692
* 82463: contig of 8772 bp in length
* 82464
* 82563: gap of unknown length
* 82564
* 90712: contig of 8149 bp in length
* 90713
* 90812: gap of unknown length
* 90813
* 110037: contig of 19225 bp in length
* 110038
* 110137: gap of unknown length
* 110138
* 137265: contig of 27128 bp in length
* 137266
* 137266: gap of unknown length
* 137366
* 181047: contig of 43682 bp in length.

FEATURES

source

1. .181047
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-555F1"
1. .1833

misc_feature

misc_feature /note="assembly_name:Contig13"
1934. .3870
/note="assembly_name:Contig14"
3971. .6886
/note="assembly_name:Contig15
clone_end:17
vector_side:right"
6987. .10052
/note="assembly_name:Contig16"
10153. .14183
/note="assembly_name:Contig17"
14284. .17263
/note="assembly_name:Contig18"
17364. .23188
/note="assembly_name:Contig19"
23289. .28005
/note="assembly_name:Contig20"
28106. .32605
/note="assembly_name:Contig21"
32706. .38209
/note="assembly_name:Contig22"
38310. .44789
/note="assembly_name:Contig23"
44890. .51559
/note="assembly_name:Contig24"
51660. .59185
/note="assembly_name:Contig25"
59286. .67326
/note="assembly_name:Contig26"
67427. .73591
/note="assembly_name:Contig27"
73692. .82463
/note="assembly_name:Contig28"
82564. .90712
/note="assembly_name:Contig29"
90813. .110037
/note="assembly_name:Contig30"
110138. .137265
/note="assembly_name:Contig31"
137366. .181047
/note="assembly_name:Contig32"
BASE COUNT 40954 a 47507 c 48668 g 42006 t 1912 others
ORIGIN

Query Match 1.2% Score 95; DB 2; Length 181047;
Best Local Similarity 100.0%; Pred. No. 7e-41;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4383 tactcggaggctgaggcaggagaatcgttgaaacccggaggcgaggttcagtgacg 4442
|||||
Db 154633 TACTCGGAGGCTGAGGCGAGGAGAAATCGCTTGAACCCGGAGCGGAGGTTGCAGTGAGC 154574

Qy 4443 cgagatacacactgcactccagcctggcgagaca 4477
|||||
Db 154573 CGAGATCACACACTGCACCTCCAGCCTGGCGGACA 154539

RESULT 9

AC092602 AC092602 166634 bp DNA linear HTG 19-JUL-2001
LOCUS Homo sapiens chromosome 2 clone RP11-134G21, WORKING DRAFT
DEFINITION SEQUENCE, 3 unordered pieces.

AC092602 AC092602 1 GI:14916187

AC092602 1 HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS human.

SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 166634)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone

Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-717A5; the clone sequenced to the right is RP11-439L14. Actual start of this clone is at base position 1 of RP11-513019; actual end is at base position 904 of RP11-439L14.

Data from AC096715 was used to finish this clone, AC016753.

FEATURES

source

Location/Qualifiers

1. .175694

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-513019"

/clone_lib="RPCI-11"

529..590

misc_feature

/note="similar to Sus scrofa EST BI339210 (NID:g15032493)"

misc_feature

992..1088

repeat_region

/note="similar to Sus scrofa EST BI339210 (NID:g15032493)"

misc_feature

1466..1576

repeat_region

/rpt_family="HY1"

misc_feature

1754..2082

repeat_region

/note="match to EST AL040160 (NID:g5409125)"

misc_feature

1976..2000

repeat_region

/rpt_family="(CACCC)n"

misc_feature

2002..2082

misc_feature

/note="similar to Sus scrofa EST BI339210 (NID:g15032493)"

misc_feature

2214..2455

misc_feature

/note="similar to Sus scrofa EST BI343196 (NID:g15036485)"

misc_feature

2372..2455

misc_feature

/note="match to EST AL040160 (NID:g5409125)"

repeat_region

2398..2422

misc_feature

/note="similar to Homo sapiens EST BG912620 (NID:g14293096)"

misc_feature

2611..2783

misc_feature

/note="similar to Bos taurus EST BE684902 (NID:g10072526)"

misc_feature

2625..2783

misc_feature

/note="match to EST AL040160 (NID:g5409125)"

misc_feature

2625..2783

misc_feature

/note="similar to Sus scrofa EST BI343196 (NID:g15036485)"

misc_feature

2625..2783

misc_feature

/note="similar to Sus scrofa EST BI339210 (NID:g15032493)"

misc_feature

2634..2783

misc_feature

/note="match to EST AW293024 (NID:g6699660)"

misc_feature

2753..3323

misc_feature

/note="similar to Homo sapiens EST BG912620 (NID:g14293096)"

misc_feature

3298..3599

misc_feature

/note="match to EST AW374393 (NID:g6879047)"

misc_feature

3425..3553

misc_feature

/note="match to EST AW293024 (NID:g6699660)"

misc_feature

3425..3767

misc_feature

/note="similar to Bos taurus EST BE684902 (NID:g10072526)"

misc_feature

3425..3527

misc_feature

/note="similar to Sus scrofa EST BI343196 (NID:g15036485)"

misc_feature

3425..3553

misc_feature

/note="match to EST AL040160 (NID:g5409125)"

misc_feature

3425..3475

misc_feature

/note="similar to Sus scrofa EST BI339210 (NID:g15032493)"

misc_feature

3796..3987

repeat_region

/note="match to EST AW293024 (NID:g6699660)"

misc_feature

3796..3888

repeat_region

/note="match to EST AL040160 (NID:g5409125)"

misc_feature

3946..3989

repeat_region

/rpt_family="L2"

misc_feature

4746..5130

misc_feature
4775..4910
/note="match to EST AW297169 (NID:g6703795)"
repeat_region
4930..4961
/rpt_family="GC-rich"
repeat_region
5024..5134
/rpt_family="MIR"
repeat_region
5828..5904
/rpt_family="L2"
repeat_region
6112..6287
/rpt_family="MERL-type"
misc_feature
6241..6489
/note="match to EST BI603789 (NID:g15496729)"
repeat_region
6695..6986
/rpt_family="Alu"
repeat_region
7210..7290
/rpt_family="MIR"
repeat_region
7291..7585
/rpt_family="Alu"
repeat_region
7586..7663
/rpt_family="MIR"
repeat_region
7673..7938
/rpt_family="Alu"
misc_feature
7699..7927
/note="match to EST BI603789 (NID:g15496729)"
repeat_region
7939..8310
/rpt_family="L1"
repeat_region
8311..8321
/rpt_family="MIR"
repeat_region
8322..8648
/rpt_family="Alu"
repeat_region
8439..8471
/rpt_family="A-rich"
repeat_region
8649..8692
/rpt_family="MIR"
repeat_region
8852..8872
/rpt_family="AT-rich"
repeat_region
8876..8935
/rpt_family="U7"
repeat_region
9226..9308
/rpt_family="MaLR"
repeat_region
9311..9448
/rpt_family="Alu"
repeat_region
9463..9764
/rpt_family="7SLRNA"
repeat_region
9801..9916
/rpt_family="MIR"
repeat_region
9923..9969
/rpt_family="L2"
repeat_region
10185..10223
/rpt_family="(CCTA)n"
repeat_region
10385..10515
/rpt_family="Alu"
misc_feature
10732..10744
/note="match to EST AW264166 (NID:g6640982) xq84g02.x1"
repeat_region
10732..10754
/rpt_family="(T)n"
repeat_region
10921..11209
/rpt_family="Alu"
repeat_region
11209..11280

Query Match 1.1% Score 89; DB 9; Length 175694;

Best Local Similarity 100.0%; Pred.No.1.6e-37;

Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4375 atccagctactcggaggctgagcagcagagatcgcttgaaccgggagcgaggttg 4434

Db 77624 ATCCAGCTACTCGGGAGGCTGAGCGAGAGATCGCTTGAACCGGCGGAGGTG 77565

Qy 4435 cagtgcgcgagatcacaccactgcactc 4463

Db 77564 CAGTGAGCGAGATCACACCACCTGCCTC 77536

RESULT 11

AC023040 178508 bp DNA linear HTG 03-MAY-2000
 LOCUS Homo sapiens chromosome 2 clone RP11-748N13 map 2, WORKING DRAFT
 DEFINITION
 AC023040.3 G1:7684460
 VERSION
 KEYWORDS HTGS_PHASE1: HTGS-DRAFT.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 178508)

REFERENCE

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 178508)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Becker, R., Bida, P.,
 Bonuslavsky, I., Boukhvalter, R., Brown, A., Burkett, G., Castle, A.,
 Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 DeArrellano, K., Dewar, K., Domingo, M., Doyle, M., Fennell, J.,
 Ferreira, P., Fitch, W., Forrest, C., Gage, D., Galagan, J.,
 Gardina, S., Grant, G., Hagos, C., Hann, L., Karatas, A., Klein, J.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Kaur, S., Locke, K.,
 Landers, T., Lehorzky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, J., McEwan, P., McGurk, A., McKernan, K.,
 McPheters, R., Meldrum, J., Meneses, L., Morrow, J., Naylor, T.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Sever, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA

on May 3, 2000 this sequence version replaced gi:7229912.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6729

Center clone name: 748_N13

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 164472 bases at least Q40

Consensus quality: 171343 bases at least Q40

Consensus quality: 174265 bases at least Q20

Insert size: 171000; agarose-gel

Insert size: 176308; sum-of-coverage

Quality coverage: 4.5 in 920 bases; agarose-gel

Quality coverage: 4.4 in 920 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs, the true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1348: contig of 1348 bp in length

* 1349 1448: gap of 100 bp

* 1449 2859: contig of 1411 bp in length

* 2860 2959: gap of 100 bp
 * 2960 5254: contig of 2295 bp in length
 * 5255 5354: gap of 100 bp
 * 5355 5768: contig of 414 bp in length
 * 5769 8888: gap of 100 bp
 * 8889 11519: contig of 2631 bp in length
 * 11520 11619: gap of 100 bp
 * 11620 15498: contig of 3779 bp in length
 * 15499 15498: gap of 100 bp
 * 15499 19579: contig of 4441 bp in length
 * 19580 20079: gap of 100 bp
 * 20080 23673: contig of 3594 bp in length
 * 23674 23773: gap of 100 bp
 * 23774 27287: contig of 3514 bp in length
 * 27288 27387: gap of 100 bp
 * 27388 33448: contig of 6061 bp in length
 * 33449 33548: gap of 100 bp
 * 33549 39008: contig of 5450 bp in length
 * 39009 39108: gap of 100 bp
 * 39109 44224: contig of 5116 bp in length
 * 44225 44324: gap of 100 bp
 * 44325 50232: contig of 5908 bp in length
 * 50233 50332: gap of 100 bp
 * 50333 58422: contig of 8090 bp in length
 * 58423 58522: gap of 100 bp
 * 58523 64440: contig of 5918 bp in length
 * 64441 64540: gap of 100 bp
 * 64541 72727: contig of 8287 bp in length
 * 72728 72827: gap of 100 bp
 * 72828 81027: contig of 8260 bp in length
 * 81028 81127: gap of 100 bp
 * 81128 91377: contig of 10250 bp in length
 * 91378 91477: gap of 100 bp
 * 91478 102460: contig of 10983 bp in length
 * 102461 102560: gap of 100 bp
 * 102561 117037: contig of 14477 bp in length
 * 117038 117137: gap of 100 bp
 * 117138 129181: contig of 12044 bp in length
 * 129182 129281: gap of 100 bp
 * 129282 149602: contig of 20721 bp in length
 * 149603 149702: gap of 100 bp
 * 149703 178508: contig of 28606 bp in length.

FEATURES

Source

1. 178508
 Jordanism: "Homo sapiens"
 /db_xref="taxon:9606"
 /map="2"
 /chromosome="2"
 /clone="RP11-748N13"
 /clone_lib="RP11-11 Human Male HAC"
 misc_feature
 1. 1348
 /note="assembly_fragment"
 misc_feature
 1449..2859
 /note="assembly_fragment"
 misc_feature
 2960..5254
 /note="assembly_fragment"
 misc_feature
 5355..8788
 /note="assembly_fragment"
 misc_feature
 8889..11519
 /note="assembly_fragment"
 clone_end:17
 vector_side:left
 11620..15398
 /note="assembly_fragment"
 misc_feature
 15499..19579
 /note="assembly_fragment"
 misc_feature
 20080..24673
 /note="assembly_fragment"
 misc_feature
 23774..27287
 /note="assembly_fragment"
 misc_feature
 27388..33448
 /note="assembly_fragment"
 misc_feature
 33549..39008

```
/note="assembly_fragment"
39109..44224
/note="assembly_fragment"
44325..50232
/note="assembly_fragment"
50333..58422
/note="assembly_fragment"
58523..64440
/note="assembly_fragment"
64541..72727
/note="assembly_fragment"
72828..81027
/note="assembly_fragment"
81128..91377
/note="assembly_fragment"
91478..102460
/note="assembly_fragment"
102561..117037
/note="assembly_fragment"
117138..129181
/note="assembly_fragment
clone_end:SP6
vector_side:right"
129282..149602
/note="assembly_fragment"
149703..178508
/note="assembly_fragment"
BASE COUNT 45506 a 42854 c 42878 g 45067 t 2203 others
ORIGIN

Query Match 1.1%; Score 89; DB 2; Length 178508;
Best Local Similarity 100.0%; Pred. No. 1.6e-37;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4389 ggagcctgagcagagagaatcgcttgaaacccggagcggaggttgacgtgagccgagat 4448
|||||
Db 96010 GGAGGCTGAGCAGGAGAAATCGCTTGAACCGGAGCGGAGGTTCACGTGAGCCGAGAT. 95951
|||||
Qy 4449 cacaccactgcactccagctggcgaca 4477
|||||
Db 95950 CACACCACCTGCACCCAGCCCTGGGGACA 95922
|||||

RESULT 12
AC010984
LOCUS AC010984 191540 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-510C1 from 2, complete sequence.
ACCESSION AC010984
VERSION AC010984.7 GI:15668160
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 191540)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 191540)
AUTHORS Du, F., Maupin, R. and Hawkins, M.
TITLE The sequence of Homo sapiens BAC clone RP11-510C1
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 191540)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 191540)
AUTHORS Waterston, R.H.
TITLE Direct Submission
```

```
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (19-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 191540)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 19, 2001 this sequence version replaced gi:14349340.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0510C01
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Plieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACE3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-397H17, 2000 bp overlap;
the clone sequenced to the right is AC023040. Actual start of this
clone is at base position 29090 of RP11-109E12.
Location/Qualifiers
1..191540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-510C1"
/clone_lib="RPCI-11"
55..455
/note="match to EST AA632464 (NID:92555878) np80e08.s1"
418..940
/note="similar to Homo sapiens EST AV732195
(NID:gl0849740)"
761..1010
/note="match to EST BF355538 (NID:gl1314612)"
795..1415
/note="match to EST BF107120 (NID:gl0889561)"
1028..1189
/note="match to EST BE350015 (NID:g9261868) ht07h01.x1"
1028..1385
/note="match to EST AL043002 (NID:g5422428)"
1034..1504
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
```


Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 20, 2000 this sequence version replaced gi:1188875.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7490

Center clone name: 209_E_9

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 146689 bases at least Q40

Consensus quality: 149124 bases at least Q30

Consensus quality: 149969 bases at least Q20

Insert size: 151000; agarose-fp

Insert size: 150565; sum-of-contigs

Quality coverage: 5.9 in Q20 bases; agarose-fp

Quality coverage: 5.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 . 9341: contig of 9341 bp in length
* 9342 9441: gap of 100 bp
* 9442 11373: contig of 1932 bp in length
* 11374 11473: gap of 100 bp
* 11474 47057: contig of 35584 bp in length
* 47058 47157: gap of 100 bp
* 47158 49227: contig of 2070 bp in length
* 49228 49327: gap of 100 bp
* 49328 54639: contig of 5312 bp in length
* 54640 54739: gap of 100 bp
* 54740 59242: contig of 4503 bp in length
* 59243 59342: gap of 100 bp
* 59343 65284: contig of 5942 bp in length
* 65285 65384: gap of 100 bp
* 65385 70923: contig of 5539 bp in length
* 70924 71023: gap of 100 bp
* 71024 78085: contig of 7062 bp in length
* 78086 78185: gap of 100 bp
* 78186 87419: contig of 9234 bp in length
* 87420 87519: gap of 100 bp
* 87520 97716: contig of 10197 bp in length
* 97717 97816: gap of 100 bp
* 97817 120620: contig of 22804 bp in length
* 120621 120720: gap of 100 bp
* 120721 150210: contig of 29490 bp in length
* 150211 150310: gap of 100 bp
* 150311 151865: contig of 1555 bp in length.
Location/Qualifiers
1 . 151865

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"

FEATURES

source

/map="8"
/clone="RP11-209E9"
/clone_lib="RPC1-11 Human Male BAC"
1 . 9341
/note="assembly_fragment
clone_end:SP6
vector_side:left"

misc_feature

misc_feature

9442 . 11373

/note="assembly_fragment"

misc_feature

11474 . 47057

/note="assembly_fragment"

misc_feature

47158 . 49227

/note="assembly_fragment"

misc_feature

49328 . 54639

/note="assembly_fragment"

misc_feature

54740 . 59242

/note="assembly_fragment"

misc_feature

59343 . 65284

/note="assembly_fragment"

misc_feature

65385 . 70923

/note="assembly_fragment"

misc_feature

71024 . 78085

/note="assembly_fragment"

misc_feature

120721 . 150210

/note="assembly_fragment"

misc_feature

150311 . 151865

/note="assembly_fragment"

clone_end:T7

vector_side:right"

BASE COUNT 45019 a 30571 c 30486 g 44487 t 1302 others

ORIGIN

Query Match 1.1%; Score 87; DB 2; Length 151865;

Best Local Similarity 100.0%; Pred. No. 2.1e-36;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4387 cgggaggtgagcgaggagaatcgcttgaaacccggaggcgaggttgagtcgagcgag 4446

|||||
Db 107399 CGGAGGCTGAGCGAGGAGAAATCGCTTGAACCGGAGGCGGAGTTGCAGTGACGCGAG 107458

QY 4447 atcacaccactgcactccagctgggc 4473

|||||

Db 107459 ATCACACCACTGCCTCCAGCCTGGGC 107485

RESULT 14

AL590113

LOCUS

DEFINITION

PROGRESS ***

ACCESSION

AL590113

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Wellcome Trust Sanger Institute

Submitted (24-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On Jan 19, 2002 this sequence version replaced gi:17154305.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Direct Submission

Submitted (24-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On Jan 19, 2002 this sequence version replaced gi:17154305.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA234D19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815: 66% of reads
Chemistry: Dye-terminator Big Dye; 38% of reads
Chemistry: Dye-primers-ancrsham: 61% of reads
Consensus quality: 165175 bases at least Q40
Consensus quality: 165285 bases at least Q30
Consensus quality: 165284 bases at least Q20
Insert size: 165292; sum-of-ctrls
Insert size: 165102; 3.1% error; adarose-fp
Quality coverage: 6.57x in Q20 bases; sum-of-ctrls Quality
Coverage: 6.88x in Q20 bases; adarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..165292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone-lib="RP11-11.1"
/clone="RP11-234D19"
1..165292
/note="assembly_fragment:00114"
BASE COUNT 49350 a 33980 c 3334 g 48628 t
ORIGIN
misc_feature
1..165292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone-lib="RP11-11.1"
/clone="RP11-234D19"
1..165292
/note="assembly_fragment:00114"
BASE COUNT 49350 a 33980 c 3334 g 48628 t
ORIGIN

Query Match 1.1% Score 87; DB 2; Length 165292;
Best Local Similarity 100.0%; Pred. No. 2e-36;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4191 agcctgagcagagaatcgtctgaaccgcggagcgaggttgcagtcagccagatca 4450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4493 AGGCTGAGGCGAGGAGCAATCGCTTGAACCGGAGCGAGGTTGCAGTACCGGAGATCA 45022
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4451 caccactgcactccagctgagcgaca 4477
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45023 CACCACATGCACATCCAGCTGCGGAGCA 45049
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
AL584643/c 142527 bp DNA linear PRI 19-JAN-2002
LOCUS Human RP11-535M15 sequence from clone RP11-535M15 on chromosome 9, complete
DEFINITION sequence.

Search completed: July 6, 2002, 22:11:58
Job time: 66083 sec


```
Query Match
Best Local Similarity 100.0%; Score 61; DB 2; Length 1656;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 gaagctgaagcagagagatgccttgaacccggagagcggagagatgcctgaagcagat 4448
Db 369 GGAGGCTGAGGCGAGGAGATGCTTGAACCGGGAGGAGTTCGATGAGCGGAGAT 310

QY 4449 c 4449
Db 309 C 309
```

```
RESULT 5
PCI-US93-11915-2/c
Sequence 2, Application PCI/US9311915
GENERAL INFORMATION:
APPLICANT: Kufe, Donald
TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCI/US93/11915
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/065001
FILING DATE: December 31, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/065001
FILING DATE: December 31, 1992
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1656
TYPE: nucleic acid
```

```
Query Match
Best Local Similarity 100.0%; Score 61; DB 5; Length 1656;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 gaagctgaagcagagagatgccttgaacccggagagcggagagatgcctgaagcagat 4448
Db 369 GGAGGCTGAGGCGAGGAGATGCTTGAACCGGGAGGAGTTCGATGAGCGGAGAT 310

QY 4449 c 4449
Db 309 C 309

RESULT 6
US-08-324-465-5/c
Sequence 5, Application US/08124465
GENERAL INFORMATION:
APPLICANT: Kufe, Donald
TITLE OF INVENTION: GENE TRANSCRIPTION AND
TITLE OF INVENTION: TUNING RADIATION METHODS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/324,465
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/065001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1725
TYPE: nucleic acid
STRANDEDNESS: double
TOPLOGY: linear
US-08-324-465-5
```

```
Query Match
Best Local Similarity 100.0%; Score 61; DB 1; Length 1725;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 gaagctgaagcagagagatgccttgaacccggagagcggagagatgcctgaagcagat 4448
Db 369 GGAGGCTGAGGCGAGGAGATGCTTGAACCGGGAGGAGTTCGATGAGCGGAGAT 310
```

```
QY 4449 c 4449
Db 309 c 309

RESULT 7
US-08-465-981-5/c
; Sequence 5, Application US/08465981
; Patent No. 5874415
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,981
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/324,465
; FILING DATE: October 17, 1994
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-465-981-5

Query Match 0.8%; Score 61; DB 2; Length 1725;
Best Local Similarity 100.0%; Pred. No. 6.3e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 ggagcgtgagcagagaatcgcttgaaacccggaggcgaggttcagtgagccagat 4448
|||||
Db 369 GGAGCGTGAGCAGAGAGATCGCTTGAACCCGGAGCGAGGTTCAGTGAGCCGAGAT 310
|||||

QY 4449 c 4449
Db 309 c 309

RESULT 8
PCT-US93-11915-5/c
; Sequence 5, Application PC/TUS9311915
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
```

```
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11915
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US93-11915-5

Query Match 0.8%; Score 61; DB 5; Length 1725;
Best Local Similarity 100.0%; Pred. No. 6.3e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 ggagcgtgagcagagaatcgcttgaaacccggaggcgaggttcagtgagccagat 4448
|||||
Db 369 GGAGCGTGAGCAGAGAGATCGCTTGAACCCGGAGCGAGGTTCAGTGAGCCGAGAT 310
|||||

QY 4449 c 4449
Db 309 c 309

RESULT 9
US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
```



```
? SEQ ID NO 7
? LENGTH: 72604
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: modified base
? LOCATION: all n positions
? OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

Query Match
Best Local Similarity 100.0%; Score 61; DB 4; Length 72604;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4388 qqagagctgaagcagagagaatcgcttaaccggagagcagagatctacatgaagcagaca 4447
|||||
Db 15692 gggagctgaagcagagagaatcgcttaaccggagagcagagatctacatgaagcagaca 15951
QY 4448 t 4448
Eb 15952 t 15952
```

```
RESULT 10
US-09-618-100B-4
: Sequence 3, Application US/08618100B
: Patent No. 4068976
: GENERAL INFORMATION:
: APPLICANT: Briggs, Michael R.
: APPLICANT: Awerx, Johan
: APPLICANT: de Vos, Piet
: APPLICANT: Staels, Bart
: APPLICANT: Croston, Glenn E.
: APPLICANT: Miller, Stephen G.
: TITLE OF INVENTION: MODULATORS OF ob GENE AND
: TITLE OF INVENTION: SCREENING METHODS THEREFOR
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: CITY: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 MB
: MEDIUM TYPE: Storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/618,100B
: FILING DATE: March 19, 1996
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/558,588
: FILING DATE: October 30, 1995
: APPLICATION NUMBER: 08/510,544
: FILING DATE: August 2, 1995
: APPLICATION NUMBER: 08/418,036
: FILING DATE: April 5, 1995
: APPLICATION NUMBER: 08/408,544
: FILING DATE: March 20, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 219/075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
```

```
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 10684 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? DESCRIPTION: Sequence between exon 1 and exon 2
: Patent No. 6068976
US-09-618-100B-3
```

```
Query Match
Best Local Similarity 100.0%; Score 54; DB 3; Length 10684;
Matches 58; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4389 gaagagctgaagcagagagaatcgcttaaccggagagcagagatctacatgaagcagaca 4446
|||||
Db 4657 GGAGGCTGAGGCGAGGAGATGCGTTGAAC GAGTGGGCGGATGCTTGACAGGAGGAG 4714
```

```
RESULT 11
US-09-328-111-741/C
: Sequence 741, Application US/99428111
: Patent No. 6262334
: GENERAL INFORMATION:
: APPLICANT: Endege, Wilson O.
: APPLICANT: Steinmann, Kathleen E.
: APPLICANT: Astle, Jon H.
: APPLICANT: Burgess, Christopher C.
: APPLICANT: Businell, Steven F.
: APPLICANT: Carroll III, Eddie
: APPLICANT: Catino, Theodore J.
: APPLICANT: Dertl, Adnan
: APPLICANT: Ford, Donna M.
: APPLICANT: Lewis, Marcia E.
: APPLICANT: Monahan, John E.
: APPLICANT: Schlegel, Robert
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: TITLE OF INVENTION: PRODOT-1S
: FILE REFERENCE: CUD-257 (US)
: CURRENT APPLICATION NUMBER: US/09/328,111
: CURRENT FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: US 60/088,871
: EARLIER FILING DATE: 1998-06-16
: NUMBER OF SEQ ID NOS: 850
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 741
: LENGTH: 579
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(579)
: OTHER INFORMATION: n = A,T,C or G
US-09-328-111-741
```

```
Query Match
Best Local Similarity 100.0%; Score 54; DB 4; Length 579;
Matches 54; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4382 ctactcggaagatgaagcagagagaatcgcttaaccggagagcagagatctacatgaagcagaca 4435
|||||
Db 213 CTACTCGGAGGCTGAGGCGAGGAGATGCTTGAATGCGGATGCTTGACAGGAGGAG 160
```

```
RESULT 12
US-09-242-948-3/C
: Sequence 3, Application US/99428448
: Patent No. 6252057
: GENERAL INFORMATION:
: APPLICANT: Brady, Matthew J.
```

; Printen, John A
; Saltiel, Alan R
; Warner-Lambert Company,
; (Outside USA)
; TITLE OF INVENTION: Protein Targeting to Glycogen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 201 Tabor Road
; CITY: Morris Plains
; STATE: NJ
; COUNTRY: US
; ZIP: 07950
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,948
; FILING DATE: 25-Feb-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,107
; FILING DATE: 30-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ashbrook, Charles W
; REFERENCE/DOCKET NUMBER: 5485-01-CA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313 996-5215
; TELEFAX: 313 996-1553
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4238..5176
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-948-3

Query Match 0.7%; Score 54; DB 4; Length 5789;
Best Local Similarity 100.0%; Pred. No. 3.2e-12;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4389 ggaggctgaggcaggagaatcgcttgaaacccggaggcggaggttgagtgagc 4442
|||||
Db 311 GGAGGCTGAGGCAGGAGAAATCGCTTGAAACCGGGAGGGGAGTTGCAGTGAGC 258

RESULT 13
US-09-210-748A-3/c
; Sequence 3, Application US/09210748A
; Patent No. 6335156
; GENERAL INFORMATION:
; APPLICANT: Hermeking, Heiko
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: 14-3-3 SIGMA ARREST THE CELL CYCLE
; FILE REFERENCE: 1107.77810
; CURRENT APPLICATION NUMBER: US/09/210.748A
; CURRENT FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/069,416
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7680
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-210-748A-3

Query Match 0.7%; Score 54; DB 4; Length 7680;
Best Local Similarity 100.0%; Pred. No. 3.1e-12;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4376 tcccagctactcgggagctgagcgaggagaatcgcttgaaacccggaggcggga 4429
|||||
Db 6086 TCCCAGCTACTCGGAGGCTGAGCGAGGAGGAATCGCTTGAAACCCGGAGCGGA 6033

RESULT 14
US-09-247-155-141
; Sequence 141, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 141
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..447
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 4..147
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.6999980926514
; OTHER INFORMATION: seq LLLFFGKLLVVG/VG
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 858..863
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 880..891
US-09-247-155-141

Query Match 0.7%; Score 52; DB 4; Length 891;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4375 atcccagctactcgggagctgagcgaggagaatcgcttgaaacccggaggc 4426
|||||
Db 714 atcccagctactcgggagctgagcgaggagaatcgcttgaaacccggaggc 765

RESULT 15
US-08-973-544-1
; Sequence 1, Application US/08973544
; Patent No. 6338950
; GENERAL INFORMATION:
; APPLICANT: WEISS, Elisabeth
; TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1

1 NUMBER OF SEQUENCES: 4
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: NIKALDO, MARCELSTEIN, MURRAY & GRAM LLP
4 STREET: 655 FIFTEENTH ST., NW, Suite 400, G St. Lobby
5 CITY: Washington
6 STATE: DC
7 COUNTRY: USA
8 ZIP: 20005-5701
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patent In Release #1.0, Version #1.40
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: 08/087973, 544
16 FILING DATE: 18-DEC-1997
17 CLASSIFICATION: 435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: PCT EP 96/02663
20 FILING DATE: 20-JUN-1996
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: EP 95109511.6
23 FILING DATE: 20-JUN-1995
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: EP 95112233.9
26 FILING DATE: 03-AUG-1995
27 ATTORNEY/AGENT INFORMATION:
28 NAME: KILLS, Monica Chin
29 REGISTRATION NUMBER: 46,105
30 REFERENCE/DOCKET NUMBER: P8441-7073
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (202) 638-5000
33 TELEFAX: (202) 638-4610
34 INFORMATION FOR SEQ ID NO: 1:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 5581 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: double
39 TOPOLOGY: linear
40 MOLECULE TYPE: DNA (genomic)
41 FEATURE:
42 NAME/KEY: exon
43 LOCATION: 48..162
44 FEATURE:
45 NAME/KEY: exon
46 LOCATION: 544..652
47 FEATURE:
48 NAME/KEY: exon
49 LOCATION: 1044..1162
50 FEATURE:
51 NAME/KEY: exon
52 LOCATION: 1475..1567
53 FEATURE:
54 NAME/KEY: exon
55 LOCATION: 1775..1797
56 FEATURE:
57 NAME/KEY: exon
58 LOCATION: 2425..2709
59 US-08-973-544-1

Query Match: 0.74; Score 52; PB 4; Length 5581;
Best Local Similarity 100.0%; Prod. No. 20-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 4475 atccagatctacgaagctgaagcgaataatccttcaaacacgaagac 4426
|||||
L5 4404 ATCCAGCTACTCGGAGAGCTGAGGACAGTGAATTCCTTGAATCCGGAATGC 4365

Search completed: July 6, 2002, 20:18:21
Job time: 54165 sec

GenCore version 4.5
Copyright (c) 1995 - 2000 CompuGen Ltd.

M n r p s m - nucleic search, using sw model

Run on: July 6, 2002, 12:44:21 : Search time 16495 seconds
(without alignments)
10279, 718 Million cell updates/sec

Title: US-09-761-466-4

RefSeq score: 7838

Seq score: 1 cctacacccatattatattat.....aatataaacattatattat. 7838

Sequenced table: g1001_NUC

Gapop 60.0 , Gapext 60.0

Sequenced: 21474536 seqs, 10817449327 residues

Word size : 0

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 6

Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main :*

1: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
2: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
3: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
4: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
5: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
6: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
7: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
8: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
9: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
10: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
11: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
13: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
14: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
15: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
16: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
17: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
18: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
19: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
20: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
21: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
22: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
23: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
24: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
25: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
26: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
27: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
28: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
29: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
30: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
31: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
32: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
33: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
34: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
35: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
36: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
37: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
38: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
39: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
40: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
41: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
42: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
43: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*

44: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
45: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
46: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
47: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
48: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
49: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
50: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
51: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
52: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
53: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
54: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
55: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
56: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
57: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
58: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
59: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
60: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
61: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
62: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
63: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
64: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
65: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
66: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
67: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
68: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
69: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
70: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
71: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
72: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
73: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
74: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*
75: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*

Prod. No. is the number of results returned by search. It has a score greater than or equal to the score of the result being printed, and is derived by analyzing the query score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7838	100.0	7838	1	Sequence 4, App 1
2	7838	100.0	7838	2	Sequence 4, App 1
3	1072	14.7	1072	3	Sequence 4, App 1
4	1072	14.7	1072	4	Sequence 4, App 1
5	526	6.7	1072	5	Sequence 4, App 1
6	489	6.2	522	6	Sequence 4, App 1
7	375	4.8	375	7	Sequence 4, App 1
8	370	4.7	410	8	Sequence 4, App 1
9	329	4.2	338	9	Sequence 4, App 1
10	182	2.3	182	10	Sequence 4, App 1
11	182	2.3	182	11	Sequence 4, App 1
12	174	2.2	457	12	Sequence 4, App 1
13	174	2.2	457	13	Sequence 4, App 1
14	174	2.2	457	14	Sequence 4, App 1
15	174	2.2	457	15	Sequence 4, App 1
16	174	2.2	457	16	Sequence 4, App 1
17	174	2.2	457	17	Sequence 4, App 1
18	174	2.2	457	18	Sequence 4, App 1
19	174	2.2	457	19	Sequence 4, App 1
20	174	2.2	457	20	Sequence 4, App 1
21	174	2.2	457	21	Sequence 4, App 1
22	174	2.2	457	22	Sequence 4, App 1
23	174	2.2	457	23	Sequence 4, App 1
24	174	2.2	457	24	Sequence 4, App 1
25	174	2.2	457	25	Sequence 4, App 1
26	174	2.2	457	26	Sequence 4, App 1
27	174	2.2	457	27	Sequence 4, App 1
28	174	2.2	457	28	Sequence 4, App 1
29	174	2.2	457	29	Sequence 4, App 1
30	174	2.2	457	30	Sequence 4, App 1
31	174	2.2	457	31	Sequence 4, App 1
32	174	2.2	457	32	Sequence 4, App 1
33	174	2.2	457	33	Sequence 4, App 1
34	174	2.2	457	34	Sequence 4, App 1
35	174	2.2	457	35	Sequence 4, App 1
36	174	2.2	457	36	Sequence 4, App 1
37	174	2.2	457	37	Sequence 4, App 1
38	174	2.2	457	38	Sequence 4, App 1
39	174	2.2	457	39	Sequence 4, App 1
40	174	2.2	457	40	Sequence 4, App 1
41	174	2.2	457	41	Sequence 4, App 1
42	174	2.2	457	42	Sequence 4, App 1
43	174	2.2	457	43	Sequence 4, App 1

32 77 1.0 32768 60 US-60-212-350-4 Sequence 4, Appli
33 77 1.0 32768 60 US-60-213-248-12 Sequence 12, Appl
34 77 1.0 32768 61 US-60-229-514-27 Sequence 27, Appl
c 35 77 1.0 61152 63 US-60-245-222-72 Sequence 72, Appl
c 36 77 1.0 61152 64 US-60-258-251-29 Sequence 29, Appl
c 37 77 1.0 124292 60 US-60-212-664-157 Sequence 157, App
c 38 77 1.0 169074 60 US-60-212-664-69 Sequence 69, Appl
c 39 77 1.0 396478 64 US-60-258-275-158 Sequence 158, App
c 40 76 1.0 8582 30 US-09-760-443-1839 Sequence 1839, Ap
c 41 76 1.0 8582 30 US-09-760-443-1990 Sequence 1990, Ap
c 42 76 1.0 8582 30 US-09-760-443-2106 Sequence 2106, Ap
c 43 76 1.0 169645 60 US-60-212-664-1 Sequence 1, Appli
c 44 76 1.0 169645 60 US-60-212-664-295 Sequence 295, App
c 45 76 1.0 179449 60 US-60-212-664-220 Sequence 220, App

ALIGNMENTS

RESULT 1
PCT-US01-01511-4
; Sequence 4, Application PC/TUS0101511
; GENERAL INFORMATION:
; APPLICANT: Beth Israel Deaconess Medical Center
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; TITLE OF INVENTION: and uses thereof
; FILE REFERENCE: 01948/069W02
; CURRENT APPLICATION NUMBER: PCT/US01/01511
; PRIOR FILING DATE: 2000-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(7838)
; OTHER INFORMATION: n = A,T,C or G
PCT-US01-01511-4

Query Match 100.0%; Score 7838; DB 1; Length 7838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctcgagccagaggttcaagaccagcctgggaacacatagggagagaccctctctccaca 60
Db 1 ctcgagccagaggttcaagaccagcctgggaacacatagggagagaccctctctccaca 60
QY 61 aaaaattaaaactagccaggtggtggcaaacacactgtagtcacagctactcagaag 120
Db 61 aaaaattaaaactagccaggtggtggcaaacacactgtagtcacagctactcagaag 120
QY 121 gctgaggtgggagatcactgtgacctggaaagtacagtgagcgcgtgatcaca 180
Db 121 gctgaggtgggagatcactgtgacctggaaagtacagtgagcgcgtgatcaca 180
QY 181 ccactgactccagcctggggagacagtgagaccctgtcaataataaatacaacaacaat 240
Db 181 ccactgactccagcctggggagacagtgagaccctgtcaataataaatacaacaacaat 240
QY 241 aatgattaaaataactaaaactatttatgctattttcacctgtattttgtaagatt 300
Db 241 aatgattaaaataactaaaactatttatgctattttcacctgtattttgtaagatt 300
QY 301 tttaaaatgaaattcccaaatgtctttccagagagattgttccaaattataccacatt 360
Db 301 tttaaaatgaaattcccaaatgtctttccagagagattgttccaaattataccacatt 360
QY 361 tcactcatgttcttctctgaacagcagcaatcaggaaaaactccctggagagcgagg 420

Db 361 tcactcatgttcttctctgaacagcagcaatcaggaaaaactccctggagagcgagg 420
QY 421 cttagactgagattttaaaaggggttaggcctcagctctcttccaggtttacactgtgc 480
Db 421 cttagactgagattttaaaaggggttaggcctcagctctcttccaggtttacactgtgc 480
QY 481 atgtttccaaactcaaaagattttacactctctgttgctgctctgtgtaaatgtcgc 540
Db 481 atgtttccaaactcaaaagattttacactctctgttgctgctctgtgtaaatgtcgc 540
QY 541 ccactactatgtattaaaaagggatgatgataatgaattcagccctctctgtaaaatcc 600
Db 541 ccactactatgtattaaaaagggatgatgataatgaattcagccctctctgtaaaatcc 600
QY 601 aaaggtctctattgcagtttccccatttaattgggttcattaaaaatttcttgggaagac 660
Db 601 aaaggtctctattgcagtttccccatttaattgggttcattaaaaatttcttgggaagac 660
QY 661 aaagcttttagtttaactatgagaaaaaacagcagacccctggattctgtcttcaag 720
Db 661 aaagcttttagtttaactatgagaaaaaacagcagacccctggattctgtcttcaag 720
QY 721 attttaccatgttgccagcctgttagtccagagcccaagaaaaatatccagccacagat 780
Db 721 attttaccatgttgccagcctgttagtccagagcccaagaaaaatatccagccacagat 780
QY 781 accctagatgtagactagcagtgctacaacctcaaggtcagaagtatgtcactagaccag 840
Db 781 accctagatgtagactagcagtgctacaacctcaaggtcagaagtatgtcactagaccag 840
QY 841 agccaaaaataggtgctatattcatttaagagagtataaaatgcacaacacagacgggtgc 900
Db 841 agccaaaaataggtgctatattcatttaagagagtataaaatgcacaacacagacgggtgc 900
QY 901 attattccaaataagcatataacccacaggggactctctctgaatatgcaagaactct 960
Db 901 attattccaaataagcatataacccacaggggactctctctgaatatgcaagaactct 960
QY 961 cactaatcaataagaaaaagcaaaagatttaaaacaggcacttcacaaaaaagtatt 1020
Db 961 cactaatcaataagaaaaagcaaaagatttaaaacaggcacttcacaaaaaagtatt 1020
QY 1021 caaaaaatcaataaacatttgaaaagatcctcaattcactagtattagggaaggtgaa 1080
Db 1021 caaaaaatcaataaacatttgaaaagatcctcaattcactagtattagggaaggtgaa 1080
QY 1081 ataaaaccacatgagacacccccacgccccacaggaagcgttaaaatctaaaaacatg 1140
Db 1081 ataaaaccacatgagacacccccacgccccacaggaagcgttaaaatctaaaaacatg 1140
QY 1141 taatcccgaaatgtttgcaaggatgcggagaaactgccattttttgtacactgccagtatga 1200
Db 1141 taatcccgaaatgtttgcaaggatgcggagaaactgccattttttgtacactgccagtatga 1200
QY 1201 gggtaaatctgtacaaacccaggttgaaaaacgctgagtagaagttaacttagctgatttg 1260
Db 1201 gggtaaatctgtacaaacccaggttgaaaaacgctgagtagaagttaacttagctgatttg 1260
QY 1261 tgaatatcatatgatccagcaatttctctcctagaaatttaccacaagaaatgtgtaaa 1320
Db 1261 tgaatatcatatgatccagcaatttctctcctagaaatttaccacaagaaatgtgtaaa 1320
QY 1321 catgttcacaaaagacacacgacaatctcatagagcactactatctcttaacagt 1380
Db 1321 catgttcacaaaagacacacgacaatctcatagagcactactatctcttaacagt 1380
QY 1381 caaaactggaaactacccaaatgtccatcagcagagaatggcgataaacagtagcatct 1440
Db 1381 caaaactggaaactacccaaatgtccatcagcagagaatggcgataaacagtagcatct 1440
QY 1441 tcacataatgaattttctgacagcaatgaaagtagctagctacactacacaaatgt 1500

Dd 1441 tcaataataaagattttccagcaagcaatgaaagagtagtagtacaactacaacaagatct 1500
Qy 1501 gatttaacrtccacaacatactaaagtaaaattatccagacacacaagaagtggtatafactc 1560
Dd 1501 gatrgaacctccacaacatactaaagtaaaattatccagacacacaagaagtggtatafactc 1560
Qy 1561 tattttadacatgtgaadctgaaacacagcacaacatactctctcttttttagaaagcaaaa 1620
Dd 1561 tattttadacatgtgaagctgaaacacagcacaacatactctctcttttttagaaagcaaaa 1620
Qy 1621 taatttactgccttgcgggaaacagaaactcaagagagcttagtagctactcgttaattgttc 1680
Dd 1621 taatttactgccttgcgggaaacagaaactcaagagagcttagtagctactcgttaattgttc 1680
Qy 1681 tcttctctgaactgcagctagtaggaagcagctgtatttttgcagctccagctgtttacact 1740
Dd 1681 tcttctctgaactgcagctagtaggaagcagctgtatttttgcagctccagctgtttacact 1740
Qy 1741 gaagttaaaagttcccccaaaatcagaagtggttcagcaagtgaagaacagtagcactgct 1800
Dd 1741 gaagttaaaagttcccccaaaatcagaagtggttcagcaagtgaagaacagtagcactgct 1800
Qy 1801 ggaacttgccttgcgggaaacttagggatcccatatttgcacagccacagaagcaagcaact 1860
Dd 1801 ggaacttgccttgcgggaaacttagggatcccatatttgcacagccacagaagcaagcaact 1860
Qy 1861 ttcttgccttaagtagcatctcccaagatccagatcccaagatggtttgcagagcagat 1920
Dd 1861 ttcttgccttaagtagcatctcccaagatccagatcccaagatggtttgcagagcagat 1920
Qy 1921 gaaagcaaatctccagagtagcagaaattttcccaatgccacttagtcccaactccccc 1980
Dd 1921 gaaagcaaatctccagagtagcagaaattttcccaatgccacttagtcccaactccccc 1980
Qy 1981 tctccacttctaatgaacggagatcacacttttttagaaaaagagctgtaggctccta 2040
Dd 1981 tctccacttctaatgaacggagatcacacttttttagaaaaagagctgtaggctccta 2040
Qy 2041 qacttaactctcacttaggaagagagagtagcactagagagctccactcacaccccc 2100
Dd 2041 ggcataactgtcacttaggaagagagagtagcactagagagctccactcacaccccc 2100
Qy 2101 tttagctatctccactctctgaattataaatgcccaacttctctatctccacacacaca 2160
Dd 2101 tttagctatctccactctctgaattataaatgcccaacttctctatctccacacacaca 2160
Qy 2161 tcttgttagaagaaagaaacgaactccacagagctctctctaaacaaagtgttcattcca 2220
Dd 2161 tcttgttagaagaaagaaacgaactccacagagctctctctaaacaaagtgttcattcca 2220
Qy 2221 gatttaacrtcccttgaagucccttgccttgaagagatggaagagagagcctcccccctcc 2280
Dd 2221 gatttaacrtcccttgaagucccttgccttgaagagatggaagagagagcctcccccctcc 2280
Qy 2281 aggaagatcatctccagagctaccacagactgaagtaactaaatccacagagtaaaccaaf 2340
Dd 2281 aggaagatcatctccagagctaccacagactgaagtaactaaatccacagagtaaaccaaf 2340
Qy 2341 aaacaaagaaagagagcatccaagcattggcccaagagagagagagagagagagagag 2400
Dd 2341 aggcagtagacaagggcattccaagcattggcccaaggaatggagggtaggtccagttcat 2400
Qy 2401 gttctctggttccagcatagacacaggggtccaaatgaacctcatgcaagaaacacagc 2460
Dd 2401 gttctctggttccagcatagacacaggggtccaaatgaacctcatgcaagaaacacagc 2460
Qy 2461 tagtctcccttctccacacagcaacttttggttactgataatacaaatcaactatttt 2520
Dd 2461 tagtctcccttctccacacagcaacttttggttactgataatacaaatcaactatttt 2520
Qy 2521 ttttttttttaactaaaggtcagataatcgtcaagagccacaggaagaaatagaagagcta 2580
Dd 2521 ttttttttttaactaaaggtcagataatcgtcaagagccacaggaagaaatagaagagcta 2580

Qy 2581 aaccacaacccctaaataatgaaagaaatacttcatctaaagaaacattcctaaagaaagaa 2640
Dd 2581 aaccacaacccctaaataatgaaagaaatacttcatctaaagaaacattcctaaagaaagaa 2640
Qy 2641 atatttttccctcccttactttctacagataatttttttttttttttttttttttttttttt 2700
Dd 2641 atgttttccctcccttactttctacaaatatttttttttttttttttttttttttttttt 2700
Qy 2701 aaatgaaacacacagcttcccaagagcaagctatttttttttttttttttttttttttttttt 2760
Dd 2701 aaatgaaacacagcttcccaagagcaagctatttttttttttttttttttttttttttttt 2760
Qy 2761 cagttggaagttaacacaaatgattcccaagcaaatccaaatgaaacattctatttttttttt 2820
Dd 2761 cagttggaagttaacacaaatgattcccaagcaaatccaaatgaaacattctatttttttttt 2820
Qy 2821 ccaagaatattcttccctacagctttctaaagaaagaaatcccaatgaaatatttttttttt 2880
Dd 2821 ccaagaatattcttccctacagctttctaaagaaagaaatcccaatgaaatatttttttttt 2880
Qy 2881 actgaaacacacagagagcttaactcccaatttttttttttttttttttttttttttttttt 2940
Dd 2881 actgaaacacacagagagcttaactcccaatttttttttttttttttttttttttttttttt 2940
Qy 2941 gaaagcccaagaaacaaatgaaatgagcttttttttttttttttttttttttttttttttttt 3000
Dd 2941 gaaagcccaagaaacaaatgaaatgagcttttttttttttttttttttttttttttttttttt 3000
Qy 3001 aaagacttaactcagagagctttgaaatgacaaatgaaacaaatgaaatatttttttttttttt 3060
Dd 3001 aaagacttaactcagagagctttgaaatgacaaatgaaacaaatgaaatatttttttttttttt 3060
Qy 3061 agacttgagactctgaaagagccagagagcttagagctaatgaaacttaaaacatttttttttt 3120
Dd 3061 ggcctttgagactctgaaagagccagagagcttagagctaatgaaacttaaaacatttttttttt 3120
Qy 3121 ctccctacagatccagagagcttagccacagaaatgaaatgaaatgaaatgaaatgaaatgaa 3180
Dd 3121 ctccctacagatccagagagcttagccacagaaatgaaatgaaatgaaatgaaatgaaatgaa 3180
Qy 3181 ctcttctgttaactgaaagagagagcttaggagagcttttttttttttttttttttttttttttt 3240
Dd 3181 ctcttctgttaactgaaagagagagctttgagagagcttttttttttttttttttttttttttttt 3240
Qy 3241 cagcccccag 3300
Dd 3241 cagcccccag 3300
Qy 3301 gattgcaaaactcagctcgtcagagctcccaaatgaaatgaaatgaaatgaaatgaaatgaaatg 3360
Dd 3301 gattgcaaaactcagctcgtcagagctcccaaatgaaatgaaatgaaatgaaatgaaatgaaatg 3360
Qy 3361 cttaaatccaaatgagctctctgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 3420
Dd 3361 cttaaatccaaatgagctctctgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 3420
Qy 3421 gaaatattcagctcagaaactagctctaaacaaatgaaatgaaatgaaatgaaatgaaatgaaatg 3480
Dd 3421 gaaatattcagctcagaaactagctctaaacaaatgaaatgaaatgaaatgaaatgaaatgaaatg 3480
Qy 3481 aaaaaaaagagagagagctccag 3540
Dd 3481 aaaaaaaagagagagagagctccag 3540
Qy 3541 gtttcaag 3600
Dd 3541 gtttcaag 3600
Qy 3601 cctctatctcagctccacag 3660
Dd 3601 cctctatctcagctccacag 3660

QY 3661 ccaagcgcggcgctgcctgacccgcagcgaaaggcccgagtcagtgctctaag 3720
|||||
Db 3661 ccaagcgcggcgctgcctgacccgcagcgaaaggcccgagtcagtgctctaag 3720
|||||
QY 3721 cgggtgcgtctctttgacagcgcggtttgggacaaacagcgggacagagataagg 3780
|||||
Db 3721 cgggtgcgtctctttgacagcgcggtttgggacaaacagcgggacagagataagg 3780
|||||
QY 3781 tgacataccagacagattggtgcgcgcgtgatactccctctccacagaaacgcgg 3840
|||||
Db 3781 tgacataccagacagattggtgcgcgcgtgatactccctctccacagaaacgcgg 3840
|||||
QY 3841 agctattaaagaccctatcgattactttatcttcttcctggaaagctcttcgcgagaga 3900
|||||
Db 3841 agctattaaagaccctatcgattactttatcttcttcctggaaagctcttcgcgagaga 3900
|||||
QY 3901 caaaagatgtccctgctaaagacacaaaggccacaaacagggaggtgtgcacaggcgac 3960
|||||
Db 3901 caaaagatgtccctgctaaagacacaaaggccacaaacaggggtgtgcacaggcgac 3960
|||||
QY 3961 gcacaaatcgcgcggggaaagcaaaacacactgacgcttagagtgacaaaacgtgtgt 4020
|||||
Db 3961 gcacaaatcgcgcggggaaagcaaaacacactgacgcttagagtgacaaaacgtgtgt 4020
|||||
QY 4021 gtccccagacagctccagagtcgcgcagggacgcgtggggcgcgagggcgacccacag 4080
|||||
Db 4021 gtccccagacagctccagagtcgcgcagggacgcgtggggcgcgagggcgacccacag 4080
|||||
QY 4081 tatggtctctgtccctggaaagttttttcacogtatgcgcgtaaacacagcacac 4140
|||||
Db 4081 tatggtctctgtccctggaaagttttttcacogtatgcgcgtaaacacagcacac 4140
|||||
QY 4141 acagagaaagtgcacttaggcgcgtgtgtgtacccggtgtgttttagcgaatt 4200
|||||
Db 4141 acagagaaagtgcacttaggcgcgtgtgtgtacccggtgtgttttagcgaatt 4200
|||||
QY 4201 taaagcacatcagggcgccatgctcacgcgtgtaatccccagacatttagagggcc 4260
|||||
Db 4201 taaagcacatcagggcgccatgctcacgcgtgtaatccccagacatttagagggcc 4260
|||||
QY 4261 gagcgggcgatcacctgaggtcggagtttcgacacacagcctggcacaatggtgaaac 4320
|||||
Db 4261 gagcgggcgatcacctgaggtcggagtttcgacacacagcctggcacaatggtgaaac 4320
|||||
QY 4321 cctgtctctacaaaaatacaaaaatagccggcgatggtgatgcgtcgtgtatccca 4380
|||||
Db 4321 cctgtctctacaaaaatacaaaaatagccggcgatggtgatgcgtcgtgtatccca 4380
|||||
QY 4381 gctactcggaggtcggcaggagaatcgcttgaacccgggagcgagggttcagtgga 4440
|||||
Db 4381 gctactcggaggtcggcaggagaatcgcttgaacccgggagcgagggttcagtgga 4440
|||||
QY 4441 gccgagatcacaccactgcactcagcctggcgacaaagcgaattccgtctaaaaaa 4500
|||||
Db 4441 gccgagatcacaccactgcactcagcctggcgacaaagcgaattccgtctaaaaaa 4500
|||||
QY 4501 ataaaaataaaataaataaataaagcccatcaactcaactcaaaagcgttactggtg 4560
|||||
Db 4501 ataaaaataaaataaataaataaagcccatcaactcaactcaaaagcgttactggtg 4560
|||||
QY 4561 gttgtaagtatccatgacacaggtctotaaaaatgaaacgctcatgtgtctcttttaa 4620
|||||
Db 4561 gttgtaagtatccatgacacaggtctotaaaaatgaaacgctcatgtgtgtctcttttaa 4620
|||||
QY 4621 gggcttgaagtctgcaactgtcatgtgtacacttaagtatggatgtgtcaacagcac 4680
|||||
Db 4621 gggcttgaagtctgcaactgtcatgtgtgtacacttaagtatggatgtgtcaacagcac 4680
|||||
QY 4681 cctttctagcgcgtcttctgtctgtaattccccgcatttcgcgaatttctgtggagcg 4740
|||||
Db 4681 cctttctagcgcgtcttctgtctgtaattccccgcatttcgcgaatttctgtggagcg 4740
|||||
QY 4741 cagaacgcccccggaagcgcgctgtgatcccgactttgtctccggtatcgcgcagct 4800
|||||

Db 4741 cagaacgcccccggaagggcggtgtgatcccgactttgtctcgggtatcgcgcagct 4800
|||||
QY 4801 tgttgctcctcgggtcccccgtgccatccccccggagagctctccacacacccgctgc 4860
|||||
Db 4801 tgttgctcctcgggtcccccgtgccatccccccggagagctctccacacacccgctgc 4860
|||||
QY 4861 gccgaattatacagacagtgaatgggttttttgggtgtgtgtgcaacacaaatttgt 4920
|||||
Db 4861 gccgaattatacagacagtgaatgggttttttgggtgtgtgtgcaacacaaatttgt 4920
|||||
QY 4921 cagctgtgttcaaatgcgtcccgcggttgaaaaacttggctgcggttaacgcacag 4980
|||||
Db 4921 cagctgtgttcaaatgcgtcccgcggttgaaaaacttggctgcggttaacgcacag 4980
|||||
QY 4981 caggttgagggacagccccgaaagaaagagagcgaggagggaaaagcgcgacccct 5040
|||||
Db 4981 caggttgagggacagccccgaaagaaagagagcgaggagggaaaagcgcgacccct 5040
|||||
QY 5041 agggccgctggccagccgtttccagcatcaattccagcactgagccgcgcagcagcacac 5100
|||||
Db 5041 agggccgctggccagccgtttccagcatcaattccagcactgagccgcgcagcagcacac 5100
|||||
QY 5101 gggctgggggtcccggaagtccggccagccgggttttggccagagccgcgcgaggtgc 5160
|||||
Db 5101 gggctgggggtcccggaagtccggccagccgggttttggccagagccgcgcgaggtgc 5160
|||||
QY 5161 ccggtgtgtagtgagactcttcacctctccggggagcgcgccagacacacccacc 5220
|||||
Db 5161 ccggtgtgtagtgagactcttcacctctccggggagcgcgccagacacacccacc 5220
|||||
QY 5221 cgcaagcgtcgctcgcccggtgttcccccgcgcgcgcgcacaaaaacagcgcgagtt 5280
|||||
Db 5221 cgcaagcgtcgctcgcccggtgttcccccgcgcgcgcgcacaaaaacagcgcgagtt 5280
|||||
QY 5281 cgccagctcttttcccaaacctgaacccgccaagccgaaggttcttccaaagtgcggt 5340
|||||
Db 5281 cgccagctcttttcccaaacctgaacccgccaagccgaaggttcttccaaagtgcggt 5340
|||||
QY 5341 tcccgggcttcacaccccgcgcgagcgcaaacccagcccgacacacacatttccctc 5400
|||||
Db 5341 tcccgggcttcacaccccgcgcgagcgcaaacccagcccgacacacacatttccctc 5400
|||||
QY 5401 ttcactgtatctgagctgttccatctgaactgaatgtcaactgatttccagctgtg 5460
|||||
Db 5401 ttcactgtatctgagctgttccatctgaactgaatgtcaactgatttccagctgtg 5460
|||||
QY 5461 acctccagcagcgtcccgaggaactgttccagcgtctcagctctctcgcctctcc 5520
|||||
Db 5461 acctccagcagcgtcccgaggaactgttccagcgtctcagctctctcgcctctcc 5520
|||||
QY 5521 gccggttttggctgaagcgttttcagcgcgtcgggcgagaaaggggtgggagtgccagcc 5580
|||||
Db 5521 gccggttttggctgaagcgttttcagcgcgtcgggcgagaaaggggtgggagtgccagcc 5580
|||||
QY 5581 accagccccagccccagaaagagagcaaatcaacgcgaaaggacacccggaagtc 5640
|||||
Db 5581 accagccccagccccagaaagagagcaaatcaacgcgaaaggacacccggaagtc 5640
|||||
QY 5641 tgaagcgaactccctcggatcctcggaatccgagcgcaaaccttaacacttagttgaaagc 5700
|||||
Db 5641 tgaagcgaactccctcggatcctcggaatccgagcgcaaaccttaacacttagttgaaagc 5700
|||||
QY 5701 ggatcatactcaactaatccagacaaaatttcgggttgggaaacatactccccagagcctaa 5760
|||||
Db 5701 ggatcatactcaactaatccagacaaaatttcgggttgggaaacatactccccagagcctaa 5760
|||||
QY 5761 gaaaactgacttacaacaaaaacaaactgaacaggaacaaaatgcgaagaggtttgtgaaa 5820
|||||
Db 5761 gaaaactgacttacaacaaaaacaaactgaacaggaacaaaatgcgaagaggtttgtgaaa 5820
|||||
QY 5821 cgtaatgtctcagaaaatgtgtatatatatatacactcctataatatgttttaaatgtg 5880
|||||


```

1b 5821 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 5889
0Y 5881 caaaaaaaatctctaaagaaataatattttaaaccacaaatgaaacttgaagaaat 5949
1b 5881 caaaaaaaatctctaaagaaataatattttaaaccacaaatgaaacttgaagaaat 5949
0Y 5941 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6009
1b 5941 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6009
0Y 6001 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6060
1b 6001 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6060
0Y 6061 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6120
1b 6061 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6120
0Y 6121 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6180
1b 6121 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6180
0Y 6181 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6240
1b 6181 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6240
0Y 6241 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6309
1b 6241 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6309
0Y 6301 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6360
1b 6301 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6360
0Y 6361 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6420
1b 6361 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6420
0Y 6421 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6480
1b 6421 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6480
0Y 6481 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6540
1b 6481 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6540
0Y 6541 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6600
1b 6541 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6600
0Y 6601 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6660
1b 6601 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6660
0Y 6661 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6720
1b 6661 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6720
0Y 6721 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6780
1b 6721 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6780
0Y 6781 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6840
1b 6781 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6840
0Y 6841 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6900
1b 6841 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6900
0Y 6901 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6960
1b 6901 aaataatgactctcaaaaataatgataatataatataatataatgatttataat 6960

```

RESULT 2

US-09-761-466-4

: Sequence 4: Application US-09-761-466

: GENERAL INFORMATION:

: ASPIRANT: Lewis, Ike W.

: ASPIRANT: Lewis, Ike W.

: TITLE OF INVENTION: Candidate's 11 Sequence Information File

: TITLE OF INVENTION: and Uses thereof

: FILE REFERENCE: 01948/US-09-761-466

: CURRENT APPLICATION NUMBER: US-09-761-466

: CURRENT FILING DATE: 2001-07-16

; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(7838)
; OTHER INFORMATION: n = A,T,C or G
US-09-761-466-4

Query Match 100.0%; Score 7838; DB 30; Length 7838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ctcgagcccgaggttcaagaccagcctgggaacatagggagacccctctctccaca 60
Dbb 1 ctcgagcccgaggttcaagaccagcctgggaacatagggagacccctctctccaca 60
Qy 61 aaaaatttaaaactagccaggtgtgtgtggcaaacacactgtagtcgccagctactcagaag 120
Dbb 61 aaaaatttaaaactagccaggtgtgtgtggcaaacacactgtagtcgccagctactcagaag 120
Qy 121 gctcaggtggagagatcacttgacgtggaagtagaggtacagctgagccgtgatccaca 180
Dbb 121 gctcaggtggagagatcacttgacgtggaagtagaggtacagctgagccgtgatccaca 180
Qy 181 ccactgcactccagcctgggagacagatgagacccctgtcaaaataaaacaaacaat 240
Dbb 181 ccactgcactccagcctgggagacagatgagacccctgtcaaaataaaacaaacaat 240
Qy 241 aatgattaaaactaaaactaaatttatgctattttcaacctgtattttgtaaagatt 300
Dbb 241 aatgattaaaactaaaactaaatttatgctattttcaacctgtattttgtaaagatt 300
Qy 301 tttaaaatgaaaatcccaaatgtttccagaaggatgttcaaaatataccacact 360
Dbb 301 tttaaaatgaaaatcccaaatgtttccagaaggatgttcaaaatataccacact 360
Qy 361 tcaactcattctctctgaacagcagcaatcaggaaaaactccctggagagcgagg 420
Dbb 361 tcaactcattctctctgaacagcagcaatcaggaaaaactccctggagagcgagg 420
Qy 421 cttgagactgagattttaaaaggggtagcctcagctccttccaggtttacactgtgc 480
Dbb 421 cttgagactgagattttaaaaggggtagcctcagctccttccaggtttacactgtgc 480
Qy 481 atgtttccaaactcaagaatttacactcttctgtgttcattgtctgttaaagatctgac 540
Dbb 481 atgtttccaaactcaagaatttacactcttctgtgttcattgtctgttaaagatctgac 540
Qy 541 ccactactatgtatttaaaagggatgcataatgaatttcagccctctctgtaaaatcc 600
Dbb 541 ccactactatgtatttaaaagggatgcataatgaatttcagccctctctgtaaaatcc 600
Qy 601 aaagggctcattgagtttcccccatttaattgggtcatlaaaatatcttgggaaggac 660
Dbb 601 aaagggctcattgagtttcccccatttaattgggtcatlaaaatatcttgggaaggac 660
Qy 661 aaagcttttagttaactatgaaaaaacagcagacaccagccctggattctgtcttcaag 720
Dbb 661 aaagcttttagttaactatgaaaaaacagcagacaccagccctggattctgtcttcaag 720
Qy 721 attttaccatgttggcagccctgtgtccagagcccaagaataatccagccacagat 780
Dbb 721 attttaccatgttggcagccctgtgtccagagcccaagaataatccagccacagat 780
Qy 781 accctagatgtagactagcagtgtacaacctcaagggtcagaagtatgtcactagaccag 840
Dbb 781 accctagatgtagactagcagtgtacaacctcaagggtcagaagtatgtcactagaccag 840

Db 781 accctagatgtagactagcagtgtacaacctcaagggtcagaagtatgtcactagaccag 840
Qy 841 agccaaaaataggtgtctatatcatcattaaagagagtaaaaaatgcaaacacagacaggtgac 900
Dbb 841 agccaaaaataggtgtctatatcatcattaaagagagtaaaaaatgcaaacacagacaggtgac 900
Qy 901 attattcaataaagcataataacccacaggggactcctctatctgaatatgtcaaaagaactct 960
Dbb 901 attattcaataaagcataataacccacaggggactcctctatctgaatatgtcaaaagaactct 960
Qy 961 cactaatcaataaagaaagcacaagatttaaacagagcacttccaaaaaaagtattatt 1020
Dbb 961 cactaatcaataaagaaagcacaagatttaaacagagcacttccaaaaaaagtattatt 1020
Qy 1021 caaaaaatacaataaacatttgaagaatcctcaattcactagttattagggaaaggtgaa 1080
Dbb 1021 caaaaaatacaataaacatttgaagaatcctcaattcactagttattagggaaaggtgaa 1080
Qy 1081 ataaacacacatgagacaccccccccgccccaccgagcgcttaaaatcttaaacatg 1140
Dbb 1081 ataaacacacatgagacaccccccccgccccaccgagcgcttaaaatcttaaacatg 1140
Qy 1141 taataccgaattgttgcgaaggatgcgagaaactgccattttttgtacactgccagtatga 1200
Dbb 1141 taataccgaattgttgcgaaggatgcgagaaactgccattttttgtacactgccagtatga 1200
Qy 1201 gggtaaatctgtacaaccaggttggaaaacgctgagtagaattactctagctggatttg 1260
Dbb 1201 gggtaaatctgtacaaccaggttggaaaacgctgagtagaattactctagctggatttg 1260
Qy 1261 tgaatatctatgatccagcaatttctactcctagaaatttaccacacagaaaatgtgtaaa 1320
Dbb 1261 tgaatatctatgatccagcaatttctactcctagaaatttaccacacagaaaatgtgtaaa 1320
Qy 1321 catgttccacaaaagacacacgaacaaattcatagaggcaactcactatctctaacagt 1380
Dbb 1321 catgttccacaaaagacacacgaacaaattcatagaggcaactcactatctctaacagt 1380
Qy 1381 caaaaactggaaactaccacaaatgtccatcagcagagaaatggcgataaacagtagcatct 1440
Dbb 1381 caaaaactggaaactaccacaaatgtccatcagcagagaaatggcgataaacagtagcatct 1440
Qy 1441 tcacataatgaatgtttcgacagcaatgaaaagttagctagctacaaactacaacaatgt 1500
Dbb 1441 tcacataatgaatgtttcgacagcaatgaaaagttagctagctacaaactacaacaatgt 1500
Qy 1501 gattgaacctcaaaaacataatactaatgtaaaattatcagacacaaagagtgatatactg 1560
Dbb 1501 gattgaacctcaaaaacataatactaatgtaaaattatcagacacaaagagtgatatactg 1560
Qy 1561 tatttagatcacatgtgaagtctgaaacacagggcaaaaactattctgtttagaagtcagaa 1620
Dbb 1561 tatttagatcacatgtgaagtctgaaacacagggcaaaaactattctgtttagaagtcagaa 1620
Qy 1621 tagttactccctgcgggaaacagacactcaagagggcttagtagctactcgttaagtctc 1680
Dbb 1621 tagttactccctgcgggaaacagacactcaagagggcttagtagctactcgttaagtctc 1680
Qy 1681 tgcctcctgaactgcagtctagtgaggcagctgtattttgtgcagtcctgtgttacact 1740
Dbb 1681 tgcctcctgaactgcagtctagtgaggcagctgtattttgtgcagtcctgtgttacact 1740
Qy 1741 ggagttaaaagttcccccaaaaatcagaaggtgttcagcaagtggaagcaagtagcacctgct 1800
Dbb 1741 ggagttaaaagttcccccaaaaatcagaaggtgttcagcaagtggaagcaagtagcacctgct 1800
Qy 1801 ggaacttggctggaaacttaggggatcccaataatttgcacaggcacaagcaagccaagct 1860
Dbb 1801 ggaacttggctggaaacttaggggatcccaataatttgcacaggcacaagcaagccaagct 1860
Qy 1861 ttcttgccntaagtagcatctccaggtccaggtatccaggaatgggtttggcaggcaggaat 1920
Dbb 1861 ttcttgccntaagtagcatctccaggtccaggtatccaggaatgggtttggcaggcaggaat 1920

[illegible]

Db	5161	ccggtggttaggtgcgactcttccactctccgggggagcgccgagaccccaacc	52220
Qy	5221	cgcaagcgctgcctcgcccgcttgcctcccgcgcgccgacaaaaacagcgcgagtt	5280
Db	5221	cgcaagcgctgcctcgcccgcttgcctcccgcgcgccgacaaaaacagcgcgagtt	5280
Qy	5281	cgccagctctctttcccaaacctgaaccgccaagccgaaggtttctccaaagtcggt	5340
Db	5281	cgccagctctctttcccaaacctgaaccgccaagccgaaggtttctccaaagtcggt	5340
Qy	5341	tcccgggcttcacaccgcgcggcagcgcgaaaccgagcccccaggaacacatttcttc	5400
Db	5341	tcccgggcttcacaccgcgcggcagcgcgaaaccgagcccccaggaacacatttcttc	5400
Qy	5401	ttcactgtatctgagtcgtttccactgtactcgaatgtcaactgatttccagctgtg	5460
Db	5401	ttcactgtatctgagtcgtttccactgtactcgaatgtcaactgatttccagctgtg	5460
Qy	5461	aactccagcgagcgactccgaggaactgattccagctgtctcgattctctccgctctcc	5520
Db	5461	aactccagcgagcgactccgaggaactgattccagctgtctcgattctctccgctctcc	5520
Qy	5521	gcccgctttggctgaagcgcttgcagccgtcgggcgcaaaaggggtgggatgtgcgcagcc	5580
Db	5521	gcccgctttggctgaagcgcttgcagccgtcgggcgcaaaaggggtgggatgtgcgcagcc	5580
Qy	5581	accagcccagccagagaaaaagagacgaaattaacgcgaaagagacaccggaagtc	5640
Db	5581	accagcccagccagagaaaaagagacgaaattaacgcgaaagagacaccggaagtc	5640
Qy	5641	tgaagcgactccctcgatctctcgaaatccgagggcaaaccttaacactagtttgaagc	5700
Db	5641	tgaagcgactccctcgatctctcgaaatccgagggcaaaccttaacactagtttgaagc	5700
Qy	5701	ggatcatatccactaatccaggagacaaattcgggttggggaacatactcccagagcctaa	5760
Db	5701	ggatcatatccactaatccaggagacaaattcgggttggggaacatactcccagagcctaa	5760
Qy	5761	gaaactgacttacaacaaaaaaactgacaaggaacaaaatgcgaagagtttgtgaaa	5820
Db	5761	gaaactgacttacaacaaaaaaactgacaaggaacaaaatgcgaagagtttgtgaaa	5820
Qy	5821	cgtaatgtctcagaaaaaatgtgtatataatacatcctataatgtttttaaatttg	5880
Db	5821	cgtaatgtctcagaaaaaatgtgtatataatacatcctataatgtttttaaatttg	5880
Qy	5881	caaaaaaaagtctctaagagatataatttttaaaaccagtggcagcttgggagggagtg	5940
Db	5881	caaaaaaaagtctctaagagatataatttttaaaaccagtggcagcttgggagggagtg	5940
Qy	5941	gggattagctgagaagggagagaagaagcattttgagtgagcgtaaaattgtttgtatc	6000
Db	5941	gggattagctgagaagggagagaagaagcattttgagtgagcgtaaaattgtttgtatc	6000
Qy	6001	tggattatggttggttgttatggggtgcacatccaagtgccaagactcatcgaaactgtac	6060
Db	6001	tggattatggttggttgttatggggtgcacatccaagtgccaagactcatcgaaactgtac	6060
Qy	6061	acttttgttctagggtacatatagacctcaataaagtggattttaaacctaaataagccagg	6120
Db	6061	acttttgttctagggtacatatagacctcaataaagtggattttaaacctaaataagccagg	6120
Qy	6121	taacagcttgcctgggtggcttggggagagggcttgggacactttacattgatctccctc	6180
Db	6121	taacagcttgcctgggtggcttggggagagggcttgggacactttacattgatctccctc	6180
Qy	6181	ttaggcatgttcggtttgtttgtttgtttcttatgatagtattattttcaaaaatat	6240
Db	6181	ttaggcatgttcggtttgtttgtttgtttcttatgatagtattattttcaaaaatat	6240
Qy	6241	atcattagcagagtactgatgtaaatgttaaaaccttgttaaggaaacacacaaagcgc	6300
Db	6241	atcattagcagagtactgatgtaaatgttaaaaccttgttaaggaaacacacaaagcgc	6300


```
Query Match 4.7% Score 370; DB 25; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4749 cttccgctaaagcgagctgtatcccaacttttctccctatctcgcacacttgttgcgc 4808
Dbb 5 cttcccaaaaaggcggtgtatcccgactttgtcccgatctgcgcagcttgttgcgc 68
QY 4809 tccgattccccatccatgcgcccccggagagctctccaaacacgcgttcgcccaatt 4868
Dbb 6 tccgattccccgtgccatgcgcccccggagagctctccaaacacgcgttcgcccaatt 128
QY 4869 atacaaactaaataggttttttttttttttttttttttttttttttttttttttttt 4928
Dbb 129 atacaaactaaataggttttttttttttttttttttttttttttttttttttttttt 188
QY 4929 gtccaaatgcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4988
Dbb 189 gtccaaatgcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 248
QY 4989 aggtcacacccgaaaggaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 5048
Dbb 249 aggtcacacccgaaaggaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 308
QY 5049 tgcgcagcgtttccacacatcaattcaacactaaagccgcgcgcgcgcgcgcgcgcgcgc 5108
Dbb 309 tgcgcagcgtttccacacatcaattcaacactaaagccgcgcgcgcgcgcgcgcgcgcgc 368
QY 5109 agctcccccga 5118
Dbb 369 agctcccccga 378

RESULT 10
US-60-182-316-1254
: Sequence 1254, Application US/60182316
: GENERAL INFORMATION:
: APPLICANT: Curtis, Anne L.
: APPLICANT: Lagace, Robert E.
: APPLICANT: Klinger, Tod M.
: APPLICANT: Stave, Laura L.
: TITLE OF INVENTION: CpG Island Polynucleotides
: FILE REFERENCE: PX-0003 P
: CURRENT APPLICATION NUMBER: US/60/182,316
: CURRENT FILING DATE: 2000-02-10
: NUMBER OF SEQ. ID NOS: 14,630
: SOFTWARE: PERL Program
: SEQ. ID NO 1254
: LENGTH: 338
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/KEY: misc.feature
: OTHER INFORMATION: Incyte ID No: CpG_991027_R15_masked_fa.Contin29558
US-60-182-316-1254
```

```
Query Match 4.2% Score 329; DB 57; Length 338;
Best Local Similarity 100.0%; Pred. No. 4.1e-90;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3883 aaqctcttcgagagacacacacacacacacacacacacacacacacacacacacacac 3942
Dbb 10 aaqctcttcgagagacacacacacacacacacacacacacacacacacacacacacacac 69
QY 3943 aaqctcttcgagagacacacacacacacacacacacacacacacacacacacacacacac 4002
Dbb 70 aaqctcttcgagagacacacacacacacacacacacacacacacacacacacacacacac 129
QY 4003 aaqctcttcgagagacacacacacacacacacacacacacacacacacacacacacacac 4062
Dbb 130 aaqctcttcgagagacacacacacacacacacacacacacacacacacacacacacacac 189
```

```
QY 4063 ggcagagcgacccacacacacacacacacacacacacacacacacacacacacacacac 4122
Dbb 190 ggcagagcgacccacacacacacacacacacacacacacacacacacacacacacacac 249
QY 4123 ggcagagcgacccacacacacacacacacacacacacacacacacacacacacacacac 4182
Dbb 250 ggcagagcgacccacacacacacacacacacacacacacacacacacacacacacacac 309
QY 4183 ggcagagcgacccacacacacacacacacacacacacacacacacacacacacacacac 4242
Dbb 310 ggcagagcgacccacacacacacacacacacacacacacacacacacacacacacacac 369

RESULT 11
US-08-369-881-1374/c
: Sequence 1374, Application US/08369881
: GENERAL INFORMATION:
: APPLICANT: Wilde, Craig S.
: APPLICANT: Deleqane, Angelo M.
: APPLICANT: Mills, Pamela Kay
: APPLICANT: Pham, Migo Thi
: TITLE OF INVENTION: HUMAN CARDIAC CELL-DEVELOPMENTAL PROTEIN ENZYMES
: TITLE OF INVENTION: AND POLYPEPTIDES
: NUMBER OF SEQUENCES: 3690
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3430 Hillview Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 6.0/6.1/MS-Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/369,881
: FILING DATE: Herewith
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Barbara J. Luther
: REGISTRATION NUMBER: 33,954
: REFERENCE/DOCKET NUMBER: PC0019 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-855-0572
: INFORMATION FOR SEQ. ID NO: 1374:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 182 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: CLONE: 185540
US-08-369-881-1374
```

```
Query Match 2.5% Score 182; DB 7; Length 182;
Best Local Similarity 100.0%; Pred. No. 6.1e-47;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 taccctccacacacacacacacacacacacacacacacacacacacacacacacacac 2840
Dbb 182 taccctccacacacacacacacacacacacacacacacacacacacacacacacacac 123
QY 2841 taccctccacacacacacacacacacacacacacacacacacacacacacacacacac 2900
Dbb 122 taccctccacacacacacacacacacacacacacacacacacacacacacacacacac 63
QY 2901 taccctccacacacacacacacacacacacacacacacacacacacacacacacacac 2960
```

```
Db 62 GACTCCCATGTTATAGAGGTCATTGATGGGTTTGTGCATGAAGGCGAGGAGACTGA 3
QY 2961 ga 2962
Db 2 GA 1

RESULT 12
US-08-408-872-494/c
; Sequence 494, Application US/08408872
; GENERAL INFORMATION:
; APPLICANT: Wilde, Craig G.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Bills, Pamela K.
; APPLICANT: Pham, Mino T.
; APPLICANT: Altus, Christina M.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: CARDIAC CELL-DERIVED POLYNUCLEOTIDES AND
; NUMBER OF SEQUENCES: 1954
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3330 HILLVIEW AVENUE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,872
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PD-0032 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 852-0195
; INFORMATION FOR SEQ ID NO: 494:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; IMMEDIATE SOURCE:
; CLONE: 185540
; US-08-408-872-494

Query Match 2.3%; Score 182; DB 8; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.1e-45;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 gatcccggaagtcgaaggagcaagtcgtgtgaagccaagagaggtatctttcccta 2840
Db 182 GATCCCGGCAAGTCGGAAGGAGCAAGTCGTGTGAAGCCAAGAGGTATCTTTCCCTA 123
QY 2841 cagcttctcaagagagggggtcccggtgggtaattglagcgtggaacaccgagagct 2900
Db 122 CAGCTTCTCAAGAGAGGGGATCCCGGTGGGTAAATGTGAGCTGGAAACACCCGAGAGCT 63
QY 2901 gactcccatgtttatagaggtcattgatgggtttgtgcatggaaggcaggaggagactga 2960
Db 62 GACTCCCATGTTTATAGAGGTCATTGATGGGTTTGTGCATGAAGGCGAGGAGACTGA 3
QY 2961 ga 2962
```

```
Db 2 GA 1

RESULT 13
US-09-866-555-406/c
; Sequence 406, Application US/09866555
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Algate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Klee, Jennifer
; TITLE OF INVENTION: LUNG TUMOR-ASSOCIATED SEQUENCES AND METHODS OF USE
; FILE REFERENCE: 210121.580
; CURRENT APPLICATION NUMBER: US/09/866,555
; CURRENT FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 20487
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 406
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(457)
; OTHER INFORMATION: n = A,T,C or G
US-09-866-555-406

Query Match 2.2%; Score 174; DB 33; Length 457;
Best Local Similarity 99.4%; Pred. No. 7.9e-43;
Matches 344; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6033 coaagtgcagaactcatcgaactttgttctaggtacattagacctcaataa 6092
Db 351 CCAAGTGTCAAGACTCATCGAACTGTACACTTTTGTTC TAGGTACATTAGACCTCAATAA 292
QY 6093 agtggatttttaaccttaataagccagggtaacagctttcctgggtggctgggggagag 6152
Db 291 AGTGGATTTTAAACCTGAATAAGCCAGGTAACAGCTTTGCC TGGTGGCGGAGAGG 232
QY 6153 cttgggaacctttacattgatccctctcttaggcattgcttctgttggtttggtttc 6212
Db 231 CTTGGGACACTTTACATTGATCTCCCTCTTAGGCATGTTCTCGTTTGGTTTGTTC 172
QY 6213 ttatgatgtattattttttt-aaaaatatattcattagcagagtgactgatgataa 6271
Db 171 TTATGATGTATTATTATTTCGAAATAATATCATTTAGCAGAGTGACTGATGTAATGTAA 112
QY 6272 aaccattgttaaggaaacccaacaaagcgggaacagagacacactggtgcatcctgttaga 6331
Db 111 AACCATTTGTTAAGGAACCAACAAAGCGGGAACAGACACACTGCTGTCATCTCTGTAGA 52
QY 6332 gggataagaataagcactcgtgttccaaagctcataaaattttgg 6377
Db 51 GGGATAAGAATAAGCACTCGCTGTCCAGCTCATATAAAATATTTGG 6

RESULT 14
PCT-US01-01354-41534
; Sequence 41534, Application PC/TUS0101354
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC004PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01354
; CURRENT FILING DATE: 2001-03-17
; NUMBER OF SEQ ID NOS: 42506
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41534
; LENGTH: 8965
```


1 ; PRIOR FILING DATE: 2000-08-22
2 ; PRIOR APPLICATION NUMBER: 60/225,214
3 ; PRIOR FILING DATE: 2000-08-14
4 ; PRIOR APPLICATION NUMBER: 60/235,836
5 ; PRIOR FILING DATE: 2000-09-27
6 ; PRIOR APPLICATION NUMBER: 60/230,438
7 ; PRIOR FILING DATE: 2000-09-06
8 ; PRIOR APPLICATION NUMBER: 60/215,135
9 ; PRIOR FILING DATE: 2000-06-30
10 ; PRIOR APPLICATION NUMBER: 60/225,266
11 ; PRIOR FILING DATE: 2000-08-14
12 ; PRIOR APPLICATION NUMBER: 60/249,218
13 ; PRIOR FILING DATE: 2000-11-17
14 ; PRIOR APPLICATION NUMBER: 60/249,208
15 ; PRIOR FILING DATE: 2000-11-17
16 ; PRIOR APPLICATION NUMBER: 60/249,213
17 ; PRIOR FILING DATE: 2000-11-17
18 ; PRIOR APPLICATION NUMBER: 60/249,212
19 ; PRIOR FILING DATE: 2000-11-17
20 ; PRIOR APPLICATION NUMBER: 60/249,207
21 ; PRIOR FILING DATE: 2000-11-17
22 ; PRIOR APPLICATION NUMBER: 60/249,245
23 ; PRIOR FILING DATE: 2000-11-17
24 ; PRIOR APPLICATION NUMBER: 60/249,244
25 ; PRIOR FILING DATE: 2000-11-17
26 ; PRIOR APPLICATION NUMBER: 60/249,217
27 ; PRIOR FILING DATE: 2000-11-17
28 ; PRIOR APPLICATION NUMBER: 60/249,211
29 ; PRIOR FILING DATE: 2000-11-17
30 ; PRIOR APPLICATION NUMBER: 60/249,215
31 ; PRIOR FILING DATE: 2000-11-17
32 ; PRIOR APPLICATION NUMBER: 60/249,264
33 ; PRIOR FILING DATE: 2000-11-17
34 ; PRIOR APPLICATION NUMBER: 60/231,242
35 ; PRIOR FILING DATE: 2000-09-08
36 ; PRIOR APPLICATION NUMBER: 60/232,081
37 ; PRIOR FILING DATE: 2000-09-08
38 ; PRIOR APPLICATION NUMBER: 60/232,080
39 ; PRIOR FILING DATE: 2000-09-08
40 ; PRIOR APPLICATION NUMBER: 60/231,414
41 ; PRIOR FILING DATE: 2000-09-08
42 ; PRIOR APPLICATION NUMBER: 60/231,244
43 ; PRIOR FILING DATE: 2000-09-08
44 ; PRIOR APPLICATION NUMBER: 60/233,064
45 ; PRIOR FILING DATE: 2000-09-14
46 ; PRIOR APPLICATION NUMBER: 60/233,063
47 ; PRIOR FILING DATE: 2000-09-14
48 ; PRIOR APPLICATION NUMBER: 60/232,397
49 ; PRIOR FILING DATE: 2000-09-14
50 ; PRIOR APPLICATION NUMBER: 60/232,399
51 ; PRIOR FILING DATE: 2000-09-14
52 ; PRIOR APPLICATION NUMBER: 60/232,401
53 ; PRIOR FILING DATE: 2000-09-14
54 ; PRIOR APPLICATION NUMBER: 60/241,808
55 ; PRIOR FILING DATE: 2000-10-20
56 ; PRIOR APPLICATION NUMBER: 60/241,826
57 ; PRIOR FILING DATE: 2000-10-20
58 ; PRIOR APPLICATION NUMBER: 60/241,786
59 ; PRIOR FILING DATE: 2000-10-20
60 ; PRIOR APPLICATION NUMBER: 60/241,221
61 ; PRIOR FILING DATE: 2000-10-20
62 ; PRIOR APPLICATION NUMBER: 60/246,475
63 ; PRIOR FILING DATE: 2000-11-08
64 ; PRIOR APPLICATION NUMBER: 60/231,243
65 ; PRIOR FILING DATE: 2000-09-08
66 ; PRIOR APPLICATION NUMBER: 60/233,065
67 ; PRIOR FILING DATE: 2000-09-14

1 ; PRIOR APPLICATION NUMBER: 60/232,398
2 Query Match 1.3%; Score 98; DB 30; Length 8965;
3 Best Local Similarity 100.0%; Pred. No. 1e-19;
4 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
5 Qy 4375 atcccagctactcgggaggtgagcgagaggaatcgctgaaccccgaggcgaggttg 4434
6 |||||
7 Db 1283 atcccagctactcgggaggtgagcgagaggaatcgctgaaccccgaggcgaggttg 1342
8 |||||
9 Qy 4435 cagtgagcgagatcacaccactgcactccagcctggg 4472
10 |||||
11 Db 1343 cagtgagcgagatcacaccactgcactccagcctggg 1380

Search completed: July 7, 2002, 01:14:25
Job time: 45064 sec

normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: This is a NIH_MGC Library."

```
BASE COUNT      114 a      210 c      204 g      113 t
ORIGIN

Query Match      5.3%; Score 415; DB 10; Length 641;
Best Local Similarity 100.0%; Pred. No. 3.8e-109;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4704 gctgaatccccgcatttcgccaatttctgttgagcgcagaaacccctcccgaaagcgc 4763
Db 1 GTCTGAATCCCGCATTTGCGCAATTTGCTTGGAGCGCAGAACGCCCTCCGCGAAAGCGC 60

QY 4764 gctgctatccccgacttgcctcggtatcgcgcagctgttggcctcggtcccccgtg 4823
Db 61 GCTGCTGATCCCGACTTGTCTCCGGTATCGGCAGCTGTGTGGCTCGGGTCCCGCGTG 120

QY 4824 ccatgccccggagggtctccacagacacgcgttgcgcgaattatatacagagactgaatg 4883
Db 121 CCATGCCCGGGAGGCTCTCCACAGACACCGCTTGGCGCGAATTATACGAGACTGAATG 180

QY 4884 ggttttttgggtgtgtgtgacacacaaatttgcagctgtctgttcaaatgcgtc 4943
Db 181 GGTTTTTTGGTGTGTGTGTGCAACACAAATTTGTGAGTGTCTGTACAAATGCGGCTC 240

QY 4944 cgcggcgctggaacttgctcgttaacgcacagctttagagggcacgaccgcga 5003
Db 241 CGCCGGCGGTGGAAACTTGGCTCGGTACGCACAGCAGAGTTGGAGGGCACGACCCGGA 300

QY 5004 aggaagaaagcgcagagaggaagcgcgcacctaggcccgctggccgcgctttcc 5063
Db 301 AGGAAGCAGAGGCGAGGAGGAAAGCGGCGACCCCTAGGCCCTGGCAGCGCTTTC 360

QY 5064 agcatattcagcactgagccggccgcagcagcacagggctgggggctcccgga 5118
Db 361 AGCATCAATTCAGCACTGAGCGCGCGCAGCAGCAGCAGCGGCTGGGGGTCCCGGA 415

RESULT 2
AI393753/c
LOCUS      AI393753      461 bp      mRNA      linear      EST 04-FEB-1999
DEFINITION      tg60d12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2113175 3', mRNA sequence.
ACCESSION      AI393753
VERSION      AI393753.1 GI:4223300
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 461)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 444.
FEATURES
Location/Qualifiers
1..461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2113175"
/lab_host="Soares_NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
```

libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

```
BASE COUNT      139 a      118 c      67 g      137 t
ORIGIN

Query Match      5.1%; Score 401; DB 9; Length 461;
Best Local Similarity 99.8%; Pred. No. 5.4e-105;
Matches 451; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5926 gcttggggaggtggtggttagctgagaaaggggagaaagcatttttgaggtgacgt 5985
Db 461 GCTTGGGAGGAGTGGGATTAGCTGAGNAGGGGAGAGGACATTTTGTAGGTGACGT 402

QY 5986 aaatgtttttgattctgattatgctggtctgttatgggggtgcacatccaaagtgtcaaga 6045
Db 401 AAATGTTTTTGTATCTTGATTATGCTGGCTGTATGGGGTGCACATCCAAGTGTCAAGA 342

QY 6046 ctcatcgaaactgtaacatttcttaggtacatagacctcaataaaagtgttttaaa 6105
Db 341 CTCATCGAACTGTACACTTTTGTCTAGGTACATTAGACCTCAATAAAGTAGATTTTAA 282

QY 6106 cctaaataagccagtaaacagcttgcctggctggcgagagagcttgggacacttt 6165
Db 281 CCTAATAAGCCAGGTAAACAGCTTTGCCCTGGGTGCTGGGGGAGAGCGCTTGGGACACTTT 222

QY 6166 acattgatctccctcttaggcattgttcgttttgggttttgggttttcttatgatgtatta 6225
Db 221 ACATTGATCTCCTCTTAGGCATGTTGCTTTTGGTTTGGTTTCTTATGATGTATTA 162

QY 6226 ttattcaaaaatatatcatctagcagagtgactgatgataaaatgtaaacattgttaag 6285
Db 161 TTTATTCAAAAATATATCATTAGCAGAGTACTGATGTAATGTAAATGTAACCATTTGTTAAG 102

QY 6286 aaaccaacaagaacggggaacagagacactggtgcctctgttagagggataagaataag 6345
Db 101 AAACCAACAAGCGGGGAACAGACACTGTGTGCTCTCTGTAGAGGGATAGAATAAG 42

QY 6346 cactcgtgtccaagctcataaaatatttgg 6377
Db 41 CACTCGCTGTCCAAGCTCATAAATATTTTGG 10

RESULT 3
AA994279/c
LOCUS      AA994279      308 bp      mRNA      linear      EST 27-AUG-1998
DEFINITION      Ou05f09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1625417 3', mRNA sequence.
ACCESSION      AA994279
VERSION      AA994279.1 GI:3180824
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 308)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
```


M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

TITLE JOURNAL COMMENT

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 298.

FEATURES

source

1. .457
/organism="Homo sapiens"
/db_xref="GDB:1242571"
/db_xref="taxon:9606"
/clone="IMAGE:297649"
/clone_lib="Soares_fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p7T3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCATCTGAAGTGGGAGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHL19w."

BASE COUNT 133 a 119 c 69 g 132 t 4 others
ORIGIN

Query Match 2.2%; Score 174; DB 10; Length 457;
Best Local Similarity 99.4%; Pred. No. 4.6e-40;
Matches 344; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 6033 ccaagtgcgaagactcgcgaactgtacacatttcttaggtacattagaccctcaataa 6092
Db 351 CCAAGTGTCAAGACTCATCGAAGTGTACACTTTTGTCTAGGTACATTAGACCTCAATAA 292
QY 6093 agtggattttaaactaaataagccaggtaaacagcttgcctgggtgctggggagagg 6152
Db 291 AGTGGATTTTAAAGCTGAATAAGCCAGGTAACAGCTTTGCCGTGGTGGCTGGGGAGAGG 232
QY 6153 cttgggacactttacattgatctccctcttaggatgtctcttttggttgttggttgttc 6212
Db 231 CTTGGGACACTTTTACATTGATCTCCCTCTTAGGCATGTCGTTTGGTTGGTTTGTTC 172
QY 6213 ttatgatgtattatttttc-aaaaatatatcattagcagagtgcactgatgataatgtaa 6271
Db 171 TTATGATGATTATTATTTCGAAAAATATATCATTAGCAGAGTGACTGATGTAATGTAA 112
QY 6272 aaccattgttaagaaacccaacaaagcgggaacaaagagacactggtgcactcctgttga 6331
Db 111 AACCATTTTAAAGAAACCAACAAAGCGGGAACAAAGAGACACTGTGTCATCTCTTTAGA 52
QY 6332 gggataagaataagcactcgcgtccaaagctcataaataatttgg 6377
Db 51 GGGATAAGAATAAGCACTCGCTGTGCCAAGCTCATATAATATTTTGG 6

RESULT 6

A0262855 536 bp DNA linear GSS 24-OCT-1998
LOCUS CITBI-EI-2511A10.TR CITBI-EI Homo sapiens genomic clone 2511A10,
DEFINITION

DNA sequence.

AQ262855
AQ262855.1 GI:3789335
GSS.
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 536)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.

TITLE

Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building

JOURNAL

COMMENT

Unpublished (1998)
Other_GSSs: CITBI-EI-2511A10.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source

1. .536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2511A10"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/note="vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"

BASE COUNT 166 a 121 c 131 g 118 t
ORIGIN

Query Match 1.1%; Score 85; DB 12; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4395 tgaggcaggagaatcgcttgaaaccggagcgagggttgctgagtcgagcgatcacacc 4454
|||||

Db 108 TGAGGCAGAGAGATCGCTTGAAACCGGGAGCGGAGTTGCAGTGAGCGGAGATCACACC 167
|||||

QY 4455 actgcactccagcctggcgacaag 4479
|||||

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192
|||||

RESULT 7

LOCUS

N76766 494 bp mRNA linear EST 28-JAN-1997
YV45c05.r1 Soares fetal liver spleen INFILS Homo sapiens cDNA clone
IMAGE:245672 5' similar to contains Alu repetitive element; mRNA
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 494)

REFERENCE

AUTHORS

Chissoe, S., Dietrich, N., Dubuque, T., Faveilho, A., Gish, W., Hawkins
, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,


```
1. .376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DCAEG11"
```

BEST LOCAL SIMILARITY 100.0%;
Matches 75: Conservative 0

```
Query Match      1.0%; Score 75; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 9.5e-12;
Matches 75: Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

```
QY 4375 atccacactcggagagctgaagcagaagaatacgttgaaaccccaagcggagattg 4434
|||||
Db 285 ATCCGAGCTACTCGGAGGCTGAGCAGAGGAATGCTTTGAACCCGGGAGGTTG 226
|||||

QY 4435 caatdaacccaagatc 4449
|||||
Db 225 CAGTCAGCGGAGATC 211

RESULT 12
AW957414
LOCUS AW957414
DEFINITION EST1369604 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW957414
VERSION AW957414.1 GI:8147217
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 526)
AUTHORS Hende,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,I.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johno@tigr.org
Plate: 107
Seq primer: Reverse.
FEATURES
Source
Location/Qualifiers
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"
BASE COUNT 150 a 108 c 150 g 117 t 1 others
ORIGIN
Query Match 1.0%; Score 75; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 8.3e-12;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4375 atccacactcggagagctgaagcagaagaatacgttgaaaccccaagcggagattg 4434
|||||
Db 260 ATCCGAGCTACTCGGAGGCTGAGCAGAGGAATGCTTTGAACCCGGGAGGTTG 319
|||||

QY 4435 caatdaacccaagatc 4449
|||||
Db 120 CAGTCAGCGGAGATC 334

RESULT 13
AA527961
LOCUS AA527961
DEFINITION nh10e05.s1 NCI-CCAP_Px3 Homo sapiens cDNA clone IMAGE:954896
similar to contains Alu repetitive element; contains element MER29
MER29 repetitive element ;, mRNA sequence.
ACCESSION AA527961
VERSION AA527961.1 GI:2270030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 569)
AUTHORS Zhou,X.M., Zhang,F.P., Jiang,B.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
Wan,D.F. and Gu,J.K.
TITLE Novel human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2820)
AUTHORS Zhou,X.M., Zhang,F.P., Jiang,B.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
Wan,D.F. and Gu,J.K.
AUTHORS
```

```
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap/.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ncap@mail.nih.gov
Tissue Procurement: W. Maistrenko, Lirobab, M.L., Rodrigo Chuapal, M.D.,
Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David R. Kitzman, Ph.D.
cDNA Library Arrayed by: Genome Systems, Inc., Greg Lennon, Ph.D.
DNA Sequenced by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. database/URL at:
www-bio.lit.nih.gov/brp/imagr/imagr.html
Insert length: 564 Std Error: 9.0%
Seq primer: 40ml x fwd, El from Amersham
High quality sequence stops: 367.
FEATURES
Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:954896"
/clone_lib="NCI-IMAT:134"
/sex="Male"
/age_start="45 years old"
/age_end="45 years old"
/note="Vector: pAMBI; Size:11.5 kb; Site:2; EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 5' end of
0.060 microdissected total cells first, laterally determined to be
fully malignant prostate cancer cells. Double-stranded
cDNA was ligated to EcoRI adapters. 5 cycles of PCR
applied to the cDNA with an adapter-specific primer, and
the resulting PCR product subcloned into pAMP10 by the
BBI-cloning method (Life Tech. Inc.). Average insert
size is 600 bp. Note: Not directionally cloned. This
library was constructed by David Kitzman."
BASE COUNT 168 a 133 c 162 g 136 t
ORIGIN
Query Match 1.0%; Score 75; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 7.6e-12;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4375 atccacactcggagagctgaagcagaagaatacgttgaaaccccaagcggagattg 4434
|||||
Db 351 ATCCGAGCTACTCGGAGGCTGAGCAGAGGAATGCTTTGAACCCGGGAGGTTG 410
|||||

QY 4435 caatdaacccaagatc 4449
|||||
Db 411 CAGTCAGCGGAGATC 425

RESULT 14
AF318360
LOCUS AF318360
DEFINITION Homo sapiens pp7518 mRNA, complete cds.
ACCESSION AF318360
VERSION AF318360.1 GI:18027811
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2820)
AUTHORS Zhou,X.M., Zhang,F.P., Jiang,B.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
Wan,D.F. and Gu,J.K.
TITLE Novel human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2820)
AUTHORS Zhou,X.M., Zhang,F.P., Jiang,B.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
Wan,D.F. and Gu,J.K.
AUTHORS
```

TITLE Direct Submission
JOURNAL Submitted (02-NOV-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai 200032, P. R. China

FEATURES Location/Qualifiers
source 1. .2820 /organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1. .2820 /gene="pp7518"
CDS 1313..1621 /gene="pp7518"
/codon_start=1
/product="unknown"
/protein_id="AAL55867.1"
/db_xref="GI:18027812"
/translation="MYAGVGVSPPHWWGLETLPGLVRKSCYKTAKLLLVQVRSPTFL
ATGSIFLEDHFSMEREGGAGFMEVFLPSSGISQTLLRSTQNLHSWCAQFTGGLLLL
"

BASE COUNT 734 a 668 c 850 g 568 t
ORIGIN
Query Match 1.0%; Score 75; DB 11; Length 2820;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4376 tccagctactcggagcctgagcagcagagaatcgctgaaccgggagcgaggttgc 4435
|||||
Db 955 TCCCAGCTACTCGGAGGCTGAGCAGGAGAATCGCTTGAACCCGGGAGCGGAGTTGC 1014
|||||
Qy 4436 agtgagccgagatca 4450
|||||
Db 1015 AGTGAGCCGAGATCA 1029
|||||

RESULT 15
BE061830
LOCUS BE061830 242 bp mRNA linear EST 09-JUN-2000
DEFINITION RC1-BT0254-271199-014-g11 BT0254 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE061830
VERSION BE061830.1 GI:8406480
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 242)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC1-BT0254-271
199-014-g11&t3-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 70.

FEATURES Location/Qualifiers
source 1. .242 /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0254"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 68 a 54 c 73 g 47 t
ORIGIN
Query Match 0.9%; Score 74; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.6e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4376 tccagctactcggagcctgagcagcagagaatcgctgaaccgggagcgaggttgc 4435
|||||
Db 106 TCCCAGCTACTCGGAGGCTGAGCAGGAGAATCGCTTGAACCCGGGAGCGGAGTTGC 165
|||||
Qy 4436 agtgagccgagatc 4449
|||||
Db 166 AGTGAGCCGAGATC 179
|||||

Search completed: July 6, 2002, 15:24:07
Job time: 47166 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

GenCore - nucleic search, using sw model

Run on: July 6, 2002, 20:41:57 : Search time 1356.5 seconds
(without alignments)
8544.073 Million cell updates/sec

Title: us-09-761-466-5

Perfect score: 6751

Sequence: 1 caatttatatttaattttat.....aatttaacccgggttttt 6751

Scoring table: -ALGO_NMC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- N_Genoseq_032802.*
- 1: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1960.IAI.*
 - 2: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1961.IAI.*
 - 3: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1962.IAI.*
 - 4: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1963.IAI.*
 - 5: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1964.IAI.*
 - 6: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1965.IAI.*
 - 7: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1966.IAI.*
 - 8: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1967.IAI.*
 - 9: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1968.IAI.*
 - 10: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1969.IAI.*
 - 11: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1970.IAI.*
 - 12: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1971.IAI.*
 - 13: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1972.IAI.*
 - 14: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1973.IAI.*
 - 15: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1974.IAI.*
 - 16: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1975.IAI.*
 - 17: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1976.IAI.*
 - 18: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1977.IAI.*
 - 19: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1978.IAI.*
 - 20: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1979.IAI.*
 - 21: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA1980.IAI.*
 - 22: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA2000.IAI.*
 - 23: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA2001.IAI.*
 - 24: /SIUS1/acqdata/hold-geneseq/genoseq-emb1/NA2002.IAI.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6751	100.0	6751	22 AAS09962	Genomic DNA #2 enc
2	640	9.5	972	22 AAH44455	Human NKX2-5/Csx n
3	640	9.5	972	22 AAH48224	Heart muscle cell
4	640	9.5	972	22 AAH49590	Human NKX2-5/Csx c
5	478	7.1	476	22 AAS09963	DNA encoding hcsx/
6	412	4.6	414	21 AAB00441	Human MINT23 DNA c
7	194	2.9	1335	23 AAS78256	DNA encoding novel
8	49	0.7	729	23 AAS67506	DNA encoding novel
9	44	0.7	686	22 AA196691	Human neuroblastom

C 14	43	0.6	1525	26 AAX50674	Human secreted pro
C 11	43	0.6	1525	21 AAZ97114	Human secreted pro
C 12	43	0.6	1949	21 AAZ97034	Human secreted pro
C 13	42	0.6	138	21 AA78443	Human cancer asso
C 14	42	0.6	208	21 AAZ1443	Human gene expres
C 15	42	0.6	236	21 AA16787	Human placenta
C 16	42	0.6	51	21 AA16787	Human placenta
C 17	42	0.6	377	24 AAK5127	Human placenta
C 18	42	0.6	377	24 AAK5127	Human placenta
C 19	42	0.6	410	22 AAG3174	Human polychlor
C 20	42	0.6	724	21 AAA6126	Human polychlor
C 21	42	0.6	758	22 AA18874	Human polychlor
C 22	42	0.6	767	22 AA18875	Human polychlor
C 23	42	0.6	818	22 AA19430	Human lung blastom
C 24	42	0.6	906	21 AA78443	Human cancer asso
C 25	42	0.6	1037	21 AA78443	Human cancer asso
C 26	42	0.6	1036	22 ABA8323	Human secreted pro
C 27	42	0.6	1041	21 AA16787	Human placenta
C 28	42	0.6	1066	21 AAF2174	Human breast and o
C 29	42	0.6	1214	21 AAS9962	Human pancreatic c
C 30	42	0.6	1290	22 AAS2614	cDNA encoding for
C 31	42	0.6	1826	22 AAH44455	Human colon cancer
C 32	42	0.6	1854	22 AAS4174	cDNA encoding for
C 33	42	0.6	1977	22 AAH44455	Human colon cancer
C 34	42	0.6	1932	21 AAF2174	Human breast and o
C 35	42	0.6	3076	22 AAH4443	Human secreted pro
C 36	42	0.6	4419	22 AA15833	Human polychlor
C 37	41	0.6	186	21 AAF2174	Human breast and o
C 38	41	0.6	253	21 AAH0092	Human colon cancer
C 39	41	0.6	264	21 AAH0092	Human colon cancer
C 40	41	0.6	285	21 AAH0092	Human colon cancer
C 41	41	0.6	294	22 ABA1433	Human cell cycle
C 42	41	0.6	338	19 AAV1243	Human secreted pro
C 43	41	0.6	358	19 AAV1243	Human secreted pro
C 44	41	0.6	408	20 AAX0674	Novel nucleotide
C 45	41	0.6	432	22 AAS2944	cDNA encoding for

ALIGNMENTS

Result 1

AAS09962

10 AAS09962 standard; IMA: 6751 BP.

XX AAS09962;

XX 24-JUL-2001 (first entry)

XX become IMA #2 encoding human Csx/NKX2-5.

XX Csx/NKX2-5; cardiac enhancer; cardiac myoblast; myocyte marker; heart

XX therapeutic; heart tissue; gene therapy; human IS.

XX Homo sapiens.

XX W:200109-AZ.

XX 19-JUL-2001.

XX 16-JAN-2001: 2001W9-DS01511.

XX 14-JAN-2001: 2000IS-0176415.

XX (BETH-) BETH ISRAEL DEANESS MEDICAL CTR.

XX Lee JW, Izumo S;

XX WPI: 2001-451809/48.

XX New cardiac specific cell enhancer elements, useful for specifically
PI expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PI induction, e.g. for optimizing cardiomyocyte induction.

XX	Claim 29; Fig 4B; 66pp; English.
PS	The sequence represents the genomic sequence #2 of human Csx/NKx2.5. The
CC	nucleic acid is useful for specifically expressing a gene in a cardiac
CC	cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC	optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC	specific manner are useful for the targeted expression of genes encoding
CC	therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC	specific enhancer elements may be used for gene therapy.
XX	
SQ	Sequence 6751 BP; 1459 A; 2008 C; 1753 G; 1490 T; 41 other;
Query Match 100.0%; Score 6751; DB 22; Length 6751;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 6751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 caattctatnagtcttataaaaggaggtttttttaactcaactgacagcagga 60
DB	1 caattctatnagtcttataaaaggaggtttttttaactcaactgacagcagga 60
QY	61 ctgnaagaaaagtgaatggctctggagcttcccttaagagaccagcatggctgcc 120
DB	61 ctgnaagaaaagtgaatggctctggagcttcccttaagagaccagcatggctgcc 120
QY	121 ccaatttttttgacgtattgttcggtttttgcacctctctctctctctctctct 180
DB	121 ccaatttttttgacgtattgttcggtttttgcacctctctctctctctctctct 180
QY	181 caagacatttttgaagccaagagaatacattaccgatccacaagagcatagagtg 240
DB	181 caagacatttttgaagccaagagaatacattaccgatccacaagagcatagagtg 240
QY	241 taacagtcactgatttgttcaaataggagaggttttttttcccttttctaacc 300
DB	241 taacagtcactgatttgttcaaataggagaggttttttttcccttttctaacc 300
QY	301 ctgacccacagagctgacagttcttagaagcccccttacccgaaaaataggaaaataatccc 360
DB	301 ctgacccacagagctgacagttcttagaagcccccttacccgaaaaataggaaaataatccc 360
QY	361 ttgccaactgatttgaagggaatgctaatttttttttcttcccgagctcctcaaaa 420
DB	361 ttgccaactgatttgaagggaatgctaatttttttttcttcccgagctcctcaaaa 420
QY	421 aaaaaaaaccccttactaaaaacagggatcccgatgtagctcgatgtcccccac 480
DB	421 aaaaaaaaccccttactaaaaacagggatcccgatgtagctcgatgtcccccac 480
QY	481 taaacggttaatttcagggtccgctcacactaatctttcaaacgttcatcgagccq 540
DB	481 taaacggttaatttcagggtccgctcacactaatctttcaaacgttcatcgagccq 540
QY	541 cctggcagcagattcaacttaacagcgtccacagaccctgttccgagctctttcagc 600
DB	541 cctggcagcagattcaacttaacagcgtccacagaccctgttccgagctctttcagc 600
QY	601 gagacatttaattgaatggatgtggctgtttccagacgtcaacgctcggcgatagg 660
DB	601 gagacatttaattgaatggatgtggctgtttgacagacgtcaacgctcggcgatagg 660
QY	661 catcctctcaacagaccccccccccccgccgctcgaaaaaatcttcaaaaggcaagg 720
DB	661 catcctctcaacagaccccccccccccgccgctcgaaaaaatcttcaaaaggcaagg 720
QY	721 gggcccccaagtaggttaatttacaaccaatacgggttaacgtggcccaaaagncagcgag 780
DB	721 gggcccccaagtaggttaatttacaaccaatacgggttaacgtggcccaaaagncagcgag 780
QY	781 gaagggccgacgacctgacatgcaagtccgctcaagaagaatttgggttgaggtg 840
DB	781 gaagggccgacgacctgacatgcaagtccgctcaagaagaatttgggttgaggtg 840

QY 1921 gcaaaagrtcaagacnttactaccaaaacaaadadacacaggaagaggaaggaaggaat 1980
DB 1921 gcaaaagrtcaagacnttactaccaaaacaaadadacacaggaagaggaaggaat 1980
QY 1961 gaaagacacaaanagggaaacaaagtgaatgaatgaatgaatgaatgaatgaatgaat 2040
DB 1961 gaaagacacaaanagggaaacaaagtgaatgaatgaatgaatgaatgaatgaatgaat 2040
QY 2041 tatntctgggaagctgaacttaccagctggatgaatgaatgaatgaatgaatgaatgaat 2100
DB 2041 tatntctgggaagctgaacttaccagctggatgaatgaatgaatgaatgaatgaatgaat 2100
QY 2101 aaaaagcatctccctgggaagctgaatgaatgaatgaatgaatgaatgaatgaatgaat 2160
DB 2101 aaaaagcatctccctgggaagctgaatgaatgaatgaatgaatgaatgaatgaatgaat 2160
QY 2161 cataaacactgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2220
DB 2161 cataaacactgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2220
QY 2221 ttaacttttaatttgcacaaagacacaaagctgaatgaatgaatgaatgaatgaatgaat 2280
DB 2221 ttaacttttaatttgcacaaagacacaaagctgaatgaatgaatgaatgaatgaatgaat 2280
QY 2281 caaaacaaacaaac 2340
DB 2281 caaaacaaacaaac 2340
QY 2341 aacccatctttnagggtttttcttatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2400
DB 2341 aacccatctttnagggtttttcttatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2400
QY 2401 aataaagrttctnaaactaatttctccacacacacacacacacacacacacacacacac 2460
DB 2401 aataaagrttctnaaactaatttctccacacacacacacacacacacacacacacacac 2460
QY 2461 cccatctgaatttgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2520
DB 2461 cccatctgaatttgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2520
QY 2521 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2580
DB 2521 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2580
QY 2581 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2640
DB 2581 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2640
QY 2641 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2700
DB 2641 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2700
QY 2701 ttaagttacacaaactgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2760
DB 2701 ttaagttacacaaactgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2760
QY 2761 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2820
DB 2761 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2820
QY 2821 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2880
DB 2821 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2880
QY 2881 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2940
DB 2881 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 2940
QY 2941 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3000
DB 2941 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3000
QY 3061 aaaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3060

DB 3001 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3060
QY 3061 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3060
DB 3061 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3060
QY 3121 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3180
DB 3121 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3180
QY 3181 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3240
DB 3181 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3240
QY 3241 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3300
DB 3241 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3300
QY 3301 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3360
DB 3301 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3360
QY 3361 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3420
DB 3361 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3420
QY 3421 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3480
DB 3421 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3480
QY 3481 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3540
DB 3481 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3540
QY 3541 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3600
DB 3541 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3600
QY 3601 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3660
DB 3601 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3660
QY 3661 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3720
DB 3661 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3720
QY 3721 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3780
DB 3721 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3780
QY 3781 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3840
DB 3781 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3840
QY 3841 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3900
DB 3841 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3900
QY 3901 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3960
DB 3901 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 3960
QY 3961 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 4020
DB 3961 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 4020
QY 4021 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 4080
DB 4021 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 4080
QY 4081 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 4140
DB 4081 caaaagrttctnaaactaatttctccacacacacacacacacacacacacacacacacac 4140

XX Homo sapiens.
XX WO200026401-A1.
XX 11-MAY-2000.
XX 02-NOV-1999; 99WO-US25251.
XX 03-NOV-1998; 98US-0106925.
XX 10-MAY-1999; 99US-0309175.
XX (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX Issa J, Baylin S, Toyota M;
XX WPI; 2000-365641/31.
XX Methylated CpG island amplification (MCA) used to determine DNA
XX methylation status of CpG sites in a given locus, useful for diagnosis
XX of age-related disorders e.g. colon cancer and dementia .
XX Claim 47; Fig 6F; 100pp; English.
XX
XX The patent discloses a method known as Methylated CpG island (CGI)
XX Amplification (MCA) to identify methylated CpG sites (cytosine residues
XX 5' of neighbouring guanine). The method uses a methylation-sensitive
XX restriction endonuclease SmaI, that cleaves unmethylated CpG sites and
XX its isoschizomer xmaI, that cleaves both methylated and unmethylated
XX CGIs. It involves amplification of closely spaced methylated SmaI sites.
XX This method is used to determine the DNA methylation status of CpG sites
XX in a given locus, in normal and neoplastic cells. It can be used for
XX diagnosis of age-related disorders, associated with methylation of CGIs,
XX e.g. dementia, arteriosclerosis and diabetes mellitus, cell proliferative
XX disorders, e.g. cancer of the kidney, colon, breast, uterus, prostate and
XX lung, leukaemia, glioblastoma, astrocytoma and neuroblastoma. It is also
XX used to evaluate the cell response to a methylation modulating agent. It
XX can also be used to screen samples for the presence of hypermethylation
XX of specific genes. The present DNA sequence is the human MINT23
XX (Methylated IN Tumours) clone derived from colon cancer cell line, Caco2.
XX It is mapped to chromosome 5q34-35 and has 64% GC nucleotides. It
XX corresponds to the 3' noncoding region of the human homeobox gene, Csx.
XX
XX Sequence 414 BP; 80 A; 133 C; 133 G; 68 T; 0 other;

Query Match 4.6%; Score 312; DB 21; Length 414;
Best Local Similarity 99.5%; Pred. No. 7.8e-84;
Matches 412; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5391 cccgggtccgcagagcaactcggagtggtccacgctgcatgggtatccgagcctggtag 5450
Db 414 CCCGGGATTCGCGACAGCAACTCGGAGTGTCCAGCTGCATGGTATCCGAGCCTGGTAG 355
QY 5451 ggaaggaccgcgtggcgacacctgacagatccacctcaacagctccctgactctcg 5510
Db 354 GCAGAGGACCCGCTGGCGCGACCCCTCACCGATCCACTCAACAGCTCCCTGACTCTCG 295
QY 5511 tggggagaagggtcccaacatgacctgagtcctccctggatttggattcactctcg 5570
Db 294 GCGGAGAAAGGCGCTCCCAACATGACCTGATGCCCTGGATTGCAATTCACCTCTCG 235
QY 5571 gagacctaggaaacttttctgtccacgcgcgctgttcttctgcacgcggagagttgtg 5630
Db 234 GAGACCTAGGAACATTTTCTCTCCACGCGGCTTTGTCTTCTGCGCAGCGGAGTTTGTG 175
QY 5631 gcggcgattatgcagcgtgcaatgagtgatctcagcgtgtgtctttagctgtccccc 5690
Db 174 GCGGCGATTATGACGCTGCAATGAGTATCTCTCAGCGCTGTGTCTTAGCTGTCCCCC 115
QY 5691 aggagtgccctccgagagtgatccatgggacccccgggttggaaactgggactgagctcgga 5750
Db 114 AGGAGTGCCCTCCGAGAGTGCTCATGGGCACCCCGCTTGGAACTGGGACTGGAGCTCGGCA 55

QY 5751 cgcaggccctgagatctggcgccgccattccgcgagccagggccggcccccggg 5804
Db 54 CGCAGGCGCTGAGATCTGGCGGCCCATTTCCGCGAGCCAGGCGCGGCCGGG 1
RESULT 7
AAS78256
ID AAS78256 standard; cDNA; 1335 BP.
XX AC AAS78256;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #14060.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG14069.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX Claim 1; SEQ ID No 14060; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1335 BP; 250 A; 413 C; 410 G; 262 T; 0 other;

Query Match 2.9%; Score 194; DB 23; Length 1335;
Best Local Similarity 100.0%; Pred. No. 6.8e-49;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SOURCE		human.	
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 251)		
AUTHORS	MacDonald.M., Hucke.E., Wilkins.B.P. and McKlem.G.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire, CB10 1RO, England. E-mail contact: humquery@sanger.ac.uk		
REFERENCE	2 (bases 1 to 251)		
AUTHORS	Cross.S.H., Charlton.J.A., Nan.X. and Bird.A.P.		
TITLE	Purification of CpG islands using a methylated DNA binding column		
JOURNAL	Nat. Genet. 5 (3): 236-244 (1994)		
MEDLINE	94282070		
COMMENT	Vector: pGEM-5zf(-) Clones are available from the UK MPG Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1RO, UK. See URL: http://www.hmp.mrc.ac.uk/ for details or contact: biohelp@hmp.mrc.ac.uk		
FEATURES	Location/Qualifiers		
source	1..251		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/sex="male"		
	/dev_stage="adult"		
	/tissue_type="blood"		
	/clone_lib="CGI-1"		
	/clone="94eb"		
BASE COUNT	51 a 75 c 27 g 44 t	4 others	
ORIGIN			
	Query Match 1.2% Score 81; DB 9; Length 251;		
	Best Local Similarity 100.0%; Pred. No. le-13;		
	Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
Qy	890 aaaggaactctccagaaactaacatcaacattccagaccgcccatgacctgcgtcgtcg 949		
Db	3 AAAAGCAATCTCCACAGACTAGTACAAATTCCAGACGGCATAAGTCCTGGCTCAG 62		
Qy	950 aaactcgaggcgtgatcttcgg 970		
Db	63 AAATCGAGGGGTGAITTCGG 83		
RESULT	B		
LOCUS	AF006664	1342 bp mRNA linear ROD 04-JUL-1997	
DEFINITION	Rattus norvegicus timpan homolog (rnx-2.5) mRNA, complete cds.		
ACCESSION	AF006664		
VERSION	AF006664.1 GI:2245649		
KEYWORDS	Norway rat		
SOURCE	Rattus norvegicus		
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 1342)		
AUTHORS	Fu.Y. and Evans.S.M.		
TITLE	rnx-2.5		
JOURNAL	Dev. Genes Evol. (1997) In press		
REFERENCE	2 (bases 1 to 1342)		
AUTHORS	Fu.Y. and Evans.S.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-JUN-1997) Medicine, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0613, USA		
FEATURES	Location/Qualifiers		
source	1..1342		
	/organism="Rattus norvegicus"		
	/db_xref="taxon:10116"		
	/tissue_type="heart"		
gene	1..1342		
	/gene="rnx-2.5"		


```
XX DR WPI: 1994-341353/42.
XX PT New regulatory regions of human erythropoietin gene - used for
XX PT studying and treating diseases and for prodn. of transgenic
XX PT animal models.
XX PS Claim 1: Page 57-58; 81pp; English.
XX CC AAQ79353 shows the nt. sequence of the entire 9.3 kb genomic clone
XX CC hEpoSLH. This nucleic acid sequence includes a coding sequence, a 5'
XX CC flanking region contg. multiple regulatory elements and a 3'
XX CC flanking region contg. multiple regulatory elements. AAQ79353
XX CC consists of the last 1777 bases of AAQ79753. It corresp. to the non-
XX CC coding 3' flanking region of AAQ79753 and includes all the regulatory
XX CC elements contained therein. It extends from the 5'-most PstI site
XX CC 3' of the end of the coding sequence to a BamHI site. It comprises
XX CC 1777 bp exhibiting many stem-loop structures. It also contains TATA
XX CC boxes in forward and reverse orientation, and at least about 184
XX CC potential transcriptional regulatory elements. These elements
XX CC include TFIID, metal responsive elements, glucocorticoid
XX CC responsive elements (including GR/PR-MMTV), NF-kappa-B, AP1, AP2,
XX CC Sp1, and lymphokine responsive consensus sequences, and many
XX CC others. This region also contains two nitrogen regulatory/oxygen
XX CC sensing sequences.
XX SQ Sequence 1777 BP; 400 A; 437 C; 453 G; 487 T; 0 other.

Query Match 0.9%; Score 74; DB 15; Length 1777;
Best Local Similarity 100.0%; Pred. No. 7.2e-19;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4376 tccacgtactcggagctgaagcaggaggaatccttgaacccggagagcgaatttc 4435
DB 1670 tccacgtactcggagcgtgagcagagagaaatccttgaacccggagagcgaatttc 1611

QY 4436 agtaagccagatc 4449
DB 1610 agtgagccgagatc 1597

RESULT 13
AAQ79353/c
ID AAQ79353 standard; DNA: 9272 BP.
AC AAQ79353;
XX 05-JUN-1995 (first entry)
DE Human genomic clone hEpoSLH including erythropoietin (EPO) coding
DE sequence and 5' and 3' regulatory elements.
XX
XX Erythropoietin; erythropoiesis; red blood cell; regulatory element;
XX CAAT box; TATA box; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CAAT_signal 875..881
FT /*tag= a
FT misc_signal 1754..1759
FT /*tag= b
FT TATA_signal 1954..1959
FT /*tag= c
FT TATA_signal 2006..2010
FT /*tag= d
FT TATA_signal 2304..2309
FT /*tag= e
XX
XX W09423570-A.
XX 27-OCT-1994.
```

```
XX PF 15-APR-1994 94W0-US04141.
XX 15-APR-1993; 93US-0046295.
PR 23-JUN-1993; 93US-0082850.
XX (UUNY ) UNIV NEW YORK STATE.
XX Lee-Huang S.
XX WPI: 1994-341353/42.
XX New regulatory regions of human erythropoietin gene - used for
XX PT studying and treating diseases and for prodn. of transgenic
XX PT animal models.
XX PS Disclosure; Page 51-55; 81pp; English.
XX CC AAQ79353 shows the nt. sequence of the entire 9.3 kb genomic clone
XX CC hEpoSLH. This nucleic acid sequence includes a coding sequence, a 5'
XX CC flanking region contg. multiple regulatory elements and a 3'
XX CC flanking region contg. multiple regulatory elements.
XX SQ Sequence 9272 BP; 2055 A; 2570 C; 2426 G; 2441 T; 0 other.

Query Match 0.9%; Score 74; DB 15; Length 9272;
Best Local Similarity 100.0%; Pred. No. 6.4e-19;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4376 tccacgtactcggagcgtgagcaggaggaatccttgaacccggagagcgaatttc 4435
DB 9165 tccacgtactcggagcgtgagcaggaggaatccttgaacccggagagcgaatttc 9160

QY 4436 agtaagccagatc 4449
DB 9105 agtgagccgagatc 9094

RESULT 14
AAK87546
ID AAK87546 standard; DNA: 5211 BP.
AC AAK87546;
XX 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42358.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
XX W0200157182-A2.
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001W0-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184764.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216047.
PR 07-JUL-2000; 2000US-0216880.
```


sequence version 4.5
Copyright (c) 1995 - 2000 CompuLink, Ltd.

EM nucleic acid search, using sw model

Run date: July 6, 2002, 20:18:21 : Search time 246.41 seconds
(without alignments)
6426.493 Million cell updates/sec

Library: US-09-761-466-5
Perfect score: 6751
Sequence: 1 caattctatnagttctat.....auttaaacccagatgatt 6751

Scoring table: 0.1650_NUC

Gapop 60.0 , Gapext 60.0

Searched: 48433 seqs, 122816752 residues

Word size : 6

Total number of hits satisfying chosen parameters: 767066

Minimum hit seq length: 0

Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Database :
1: /cn2_6/ptdata/2/ina/va_cmb.seq.*
2: /cn2_6/ptdata/2/ina/vb_cmb.seq.*
3: /cn2_6/ptdata/2/ina/va_cmb.seq.*
4: /cn2_6/ptdata/2/ina/vb_cmb.seq.*
5: /cn2_6/ptdata/2/ina/va_cmb.seq.*
6: /cn2_6/ptdata/2/ina/vb_cmb.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	41	0.6	308	2	US-08-721-488-4
C 2	41	0.6	1134	3	US-09-248-445-24
C 3	41	0.6	1720	4	US-09-227-357-139
C 4	41	0.6	2209	1	US-08-514-014-1
C 5	41	0.6	2209	2	US-08-833-823-1
C 6	41	0.6	2280	3	US-08-813-150-1
C 7	41	0.6	2452	3	US-09-027-137-2
C 8	41	0.6	5962	6	538-025-5
C 9	41	0.6	5975	1	US-08-404-354R-1
C 10	41	0.6	5975	1	US-08-414-083R-1
C 11	41	0.6	5975	1	US-08-435-675R-1
C 12	41	0.6	5975	1	US-08-336-257A-3
C 13	41	0.6	5975	3	US-08-864-599-1
C 14	40	0.6	457	2	US-08-841-349-1R
C 15	40	0.6	515	3	US-08-589-02R-1
C 16	40	0.6	515	3	US-08-784-582-1
C 17	40	0.6	515	4	US-08-785-271-1
C 18	40	0.6	630	1	US-08-185-414R-1
C 19	40	0.6	790	4	US-09-461-474-14
C 20	40	0.6	806	4	US-08-955-629R-1
C 21	40	0.6	857	1	US-08-308-883-1
C 22	40	0.6	857	1	US-08-730-163-1
C 23	40	0.6	857	4	US-08-256-799-1
C 24	40	0.6	857	4	US-08-462-447-1
C 25	40	0.6	1023	1	US-08-252-966R-16
C 26	40	0.6	1056	1	US-08-157-101A-4
C 27	40	0.6	1407	2	US-08-96-022-17

C 28	40	0.6	1459	4	US-08-967-074-11
C 29	40	0.6	1459	4	US-08-967-074-11
C 30	40	0.6	1534	1	US-08-435-675R-1
C 31	40	0.6	1534	1	US-08-435-675R-1
C 32	40	0.6	1572	1	US-08-185-414R-1
C 33	40	0.6	1642	2	US-08-414-083R-1
C 34	40	0.6	1642	2	US-08-414-083R-1
C 35	40	0.6	1683	4	US-09-435-675R-1
C 36	40	0.6	1683	4	US-09-435-675R-1
C 37	40	0.6	1813	4	US-08-023-454A-2R
C 38	40	0.6	2144	4	US-08-023-454A-2R
C 39	40	0.6	2709	3	US-08-336-257A-3
C 40	40	0.6	3933	1	US-08-462-447-1
C 41	40	0.6	3933	1	US-08-462-447-1
C 42	40	0.6	3933	1	US-08-462-447-1
C 43	40	0.6	3933	1	US-08-462-447-1
C 44	40	0.6	4756	3	US-08-023-454A-2R
C 45	40	0.6	4780	3	US-08-023-454A-2R

ADDITIONS

RESULT 1
US-08-721-488-4/c
Sequence 4, Application US/08721488
Patent No. 5965388
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: LaValle, Edward
APPLICANT: Rorie, Lisa
APPLICANT: Morberg, David
APPLICANT: Lohary, Maurice
APPLICANT: Spaulding, Vicki
APPLICANT: Rowan, Michael
TITLE OF INVENTION: SEQUENTIAL PROTEINS AND ENZYMOLOGICAL
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridge Park Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER REABLE FORM:
MEDIUM TYPE: Floppy disk
COMPILER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patch II, Release 1.1, Version 1.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08721488
CLASSIFICATION: N: 4-4
ALTERNATIVE INVENTION:
NAME: Rowan, Michael
REGISTRATION NUMBER: 42724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8424
TELEFAX: (617) 876-5851
INFORMATION FOR SFO ID NO: 4
SEQUENCE CHARACTERISTICS:
LENGTH: 308 bases
TYPE: nucleic acid
STRANDEDNESS: double
TOPLOGY: linear
MODULE TYPE: cDNA
US-08-721-488-4

Query Match: 100%
Best Local Similarity: 100%
Patent No. 5965388


```
Db 5969 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCC 5929
|||||
RESULT 12
US-08-336-257A-3/c
; Sequence 3, Application US/08336257A
; Patent No. 5726035
; GENERAL INFORMATION:
; APPLICANT: Jay, Scott D
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; APPLICANT: Campbell, Kevin P.
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,257A
; FILING DATE: 07-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 54898
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 79...5700
; OTHER INFORMATION: \product="Alpha-1 subunit of animal calcium
; channel"
US-08-336-257A-3

Query Match 0.68; Score 41; DB 1; Length 5975;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2649 ttttttttttttttttttttttttttttttttttttttttttttttgc 2689
|||||
Db 5969 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCC 5929

RESULT 13
US-08-884-599-1/c
; Sequence 1, Application US/08884599
; Patent No. 6013474
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
```

```
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/884,599
; FILING DATE: 27-JUNE-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/314,083
; FILING DATE: 28-SEPT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 08-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-53191B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 79...5700
; OTHER INFORMATION:
US-08-884-599-1

Query Match 0.6%; Score 41; DB 3; Length 5975;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2649 ttttttttttttttttttttttttttttttttttttttttttttttgc 2689
|||||
Db 5969 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCC 5929

RESULT 14
US-08-841-349-18
; Sequence 18, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO44700S0
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
```


/tissue_type="glioblastoma (pooled)"
/lab_host="DH10p"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(df) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCATATCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 157 a 245 c 275 g 131 t 7 others
ORIGIN

Query Match 7.0%; Score 475; DB 9; Length 815;
Best Local Similarity 99.7%; Pred. No. 7e-84;
Matches 575; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5339 caacaacaactctgaacttcggtcgaggacttgaatcggttcagagcccgagat 5398
|||||
Db 577 CAACAACAACCTCTGTGAACCTTCGGCGTTCGGGACTTGAATCGGTTTCAGACNCCGGGAT 518
Qy 5399 tccgcagagcaactcggagtgctccacgctgcatggtattccgagcctggtagggaaggga 5458
|||||
Db 517 TCCGCAGAGCAACTCGGGAGTGTCCAGCTCATGGTATCCGAGCCTGGTAGGGAAGGGA 458
Qy 5459 ccgcgtggtcgagccctgagcagatccacactcaacagctccctgactctcgtggggaga 5518
|||||
Db 457 CCCCGGTGGCGGACCTGACGATCCACCTCAACAGCTCCCTGACTCTCGGGGGGAGA 398
Qy 5519 aggggtcccaaatgacatgacctgagtcctcgtgatttgcatcactcctcgcgagaccta 5578
|||||
Db 397 AGGGGCTCCCAACATGACCCCTGAGCTCCCTCGGATTTTGCAATCTCTCGCGAGACCTA 338
Qy 5579 ggaacttttctgtccacgcgcgtttgttcttgcacggagagatttggggcgagat 5638
Db 337 GGAACCTTTTCTGTCCACGCGCGTGTGTTCTTGGCAGCGGAGAGTTTGTGGCGGCGAT 278
Qy 5639 tatcagcgtcaatgagtatcctgagcctggtgtcttagctgtcccccagagatgc 5698
|||||
Db 277 TATCAGCGTGCATAGTATCTCTGACGCTGTGTGTTTGTAGCTGTCCCGCCAGAGTGC 218
Qy 5699 cctccgagagtcctatgggcaaccccggttggaactgggagctgagctcgggcacgcagggc 5758
Db 217 CCTCCGAGAGTCCATGGCACCCCGGTGGAACTGGGACTGAGCTCGGCGCACGAGGC 158
Qy 5759 ctgagatctggcgcgcattccgcagccagggcgcggcgcccggttgcttctctcgt 5818
Db 157 CTGAGATCTGGCGGCCCATTCGCGAGCCAGGGCGCGCGCGCGCTTTGCTATCTCG 98
Qy 5819 ccgtcgccgcgcacgcacccacccgtatttattgttttacctattgtctgaagaaatga 5878
Db 97 CCGTCGCCGCCGACGACCCACCCGCTATTTATGTTTATTACTATTGCTGTGAAGAAATGA 38
Qy 5879 cgatcccttccctattaaagagagtgggtgaccccg 5915
Db 37 CGATCCCGCTTCCCATTAAGAGAGTGCCTTGAACCCG 1

RESULT 2
AW665197/c
LOCUS
DEFINITION h301g01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2980560 3' similar to SW:HK25_HUMAN P52952 HOMBOBOX PROTEIN
NCX-2.5 ; contains Alu repetitive element;; mRNA sequence.
ACCESSION AW665197
VERSION AW665197.1 GI:7457743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 713)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 466.
FEATURES
Location/Qualifiers
1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2980560"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 145 a 205 c 236 g 125 t 2 others
ORIGIN

Query Match 6.9%; Score 466; DB 9; Length 713;
Best Local Similarity 99.6%; Pred. No. 4.6e-82;
Matches 566; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5326 ccactgcgcgcgcacaacaacttcgtgaacttcggcgctcgggacttgaatcggttc 5385
Db 584 CCAGTCGCCGCCCAACAACAACACTTCGTGAACCTTCGGGTTCGGGACTTGAATCGGTTTC 525
Qy 5386 agagccccggattccgcagagcaactcgggagtggtccaagtcgtagtataccagact 5445
Db 524 AGAGCCCGGGATTCCGAGAGCAACTCGGGAGTGTCCACGTGCATGTATCCGAGCCT 465
Qy 5446 ggtagggaagcccgctggtgcgacctgacagatccccacccaacagctccctgac 5505
Db 464 GTAGGGGAAGGACCCCGGTGCGCGGACCCCTGACCGATCCACCTCAACAGCTCCCTGAC 405
Qy 5506 tctcgtgggagaagggtcctcccaacatgacctgagtcctcctgagtttgcattcaactc 5565
Db 404 TCTCGGGGGAGAGGGGCTCCCAACATGACCTGAGTGTGATTTTGGATTTCATTCACTC 345
Qy 5566 ctgcggagacctagaaacttttctgccccgcgctttgtcttcgcacgagagagt 5625
Db 344 CTGCGGAGACCTTAGAACTTTTCTGTCCACGCGGCTTTGTCTTTCGCCACGGGAGAGT 285
Qy 5626 ttgtggcgcgattatgacagctgcaatgagtgcctcagcctgagctggtcttagctgtc 5685
Db 284 TTGTGCGCGGATTTATGACGCTGCAATGAGTGTCTCTCAGCCTGTCTTAGCTGTC 225
Qy 5686 cccccagagtgccctccgagagtcctatgggcccccccggttggaaactgggagctgc 5745
Db 224 CCCCCAGGAGTGCCCTCCGAGAGTCATCGGCACCCCGCGTTTGGAACTGGGACTGAGCTC 165
Qy 5746 gggcagcagggcctgagatctgcccgccttcacccagcagcagggcgcccgccggcg 5805
Db 164 GGGCAGCAGGGCCTGAGATCTGGCGGCCCATTCGCGATCCAGGGCCCGGCCCGCCGCG 105
Qy 5806 ctttctctctcgcgctgcgcgcgcacacacccacccctatttattgtttttacatttg 5865
Db 104 CTTTCTATCTCGCGTCGCCGCCACCGACCCACCCCTATTATGTTTTTTACCTATTG 45

Query Match 6.58; Score 48; Lp 9; Length 674;
Best Local Similarity 99.58; Pred. No. 1,40-76;
Matches 588; Conservative 0; Mismatches 0; Gaps 0;

```
Qy 5304 gggccttcccagcgagccgacactgcccgcgcgcacacaaacactctgtaacttggc 5363
|||||
Db 606 GGGCTTCCCAGCGAGCGCGCCACACTGCCGCGCCACACACACTTCTGTAACCTCGCG 547
|||||
Qy 5364 gtcggggaactgaatgcggttcagagccccgggattccgcagagcaaacctcggagtgcc 5423
|||||
Db 546 GTCGGGACTTGAATGCGGTTTCAGAGNCCGGGATTCGCGAGAGCAACTCGGGAGTGTC 487
|||||
Qy 5424 acgctgcattatccagagcctgttaggaagagaccccgctgcgagaccctaccgat 5483
|||||
Db 486 ACGCTGATGATATCCGAGCCTGTAGGGAAGGAGCCGCGTGGCGACCTCACCGAT 427
|||||
Qy 5484 cccacctcaacagctccctgactctgtgggagaaagggtccccaacatgacctgagt 5543
|||||
Db 426 CCCACCTCAACAGCTCCCTGACTCTCGGGGGGAGAGGGGTCCCAACATGACCTGAGT 367
|||||
Qy 5544 cccctgatttgcattoactcctcgcgagacactaggaactttctgtccacgcgct 5603
|||||
Db 366 CCCCTGGAATTTGCAATTCACCTCTCGCGAGACCTAGGAACCTTTTCTGTCCACGCGGT 307
|||||
Qy 5604 ttgtcttcgcacgagagatttgcggcgattatgcagctgcaatgagtgatcct 5663
|||||
Db 306 TTGTTCITGCGACGGAGAGATTGTGGCGCGGATTTATGCAAGCTGCAATGAGTGATCCT 247
|||||
Qy 5664 gcagcctggtgtcttagctgtcccccagagagtgccctccgagagtcctatggcaccccc 5723
|||||
Db 246 GCAGCCTGGTGCTTAGTGTCTCCCGCCAGGAGTGCCCTCCGAGAGTCCATGGGCACCCC 187
|||||
Qy 5724 gtttggaactgggactgagctcgggacgagggcctgagatctggcgcccatccgcg 5783
|||||
Db 186 GGTGGAACTGGGACTGAGCTCGGGGACGACGAGGGCCTGAGATCTGGCGCCCATTCGCG 127
|||||
Qy 5784 agcagggcgggcgccggcctttgctatctcgcgctgcgcgcacacacacccccc 5843
|||||
Db 126 ATCCAGGGCGGGCGCGCGGCCCTTTGCTATCTCGCCGTCGCCCGCCACCCACCCACC 67
|||||
Qy 5844 gtattatgttttaccattatgctgtgaagaatgacatcccttccatt 5894
|||||
Db 66 GTATTTATGTTTTACCTATTGCTGTGAAGAAATGACGATCCCTTCCCAT 16
|||||

RESULT 5
AW001138
LOCUS
DEFINITION
  AW001138 457 bp mRNA linear EST 27-OCT-1999
  wu24907.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
  IMAGE:2521020 3' similar to SW:HK25_HUMAN P52952 HOMEBOX PROTEIN
  NKX-2.5 ; contains PTRS.t2 TAR1 repetitive element ; , mRNA sequence.
ACCESSION
  AW001138
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 457)
  NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-r@mail.nih.gov
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Seq primer: -40UP from Gibco
  High quality sequence stop: 437.
  Location/Qualifiers
    1. .457
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2521020"
      /tissue_type="colonic mucosa from 3 patients with Crohn's"

FEATURES
  source
```

```

/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site-1: Not I; Site-2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTCTTTTCTTTTCTTTT 3'],
(double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement; from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 78 a 158 c 155 g 66 t
ORIGIN

Query Match 6.3%; Score 425; DB 9; Length 457;
Best Local Similarity 100.0%; Pred. No. 7.le-74;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4795 tctgttccccctcagagctgtgcgctgcagaaaggcgtgagctggagaagacagag 4854
|||||
Db 7 TCTGTTCCCTTCAGAGCTGTCCGCGCTGCAGAGGCGGTGGAGCTGGAAGAAGACAGAGG 66
|||||
Qy 4855 cggacaacgagagcgccccggcgacgagcgaggagaaagccgcgtgctctctcgc 4914
|||||
Db 67 CGGACAAACGCGAGCGGCCCGGGCGGACGCGGAGGAAGCGCGTCTCTCTCGC 126
|||||
Qy 4915 aggcgcaggtctatgagctggagcggtctcaagcagcagcggtacctgtcgcccccg 4974
|||||
Db 127 AGGCGCAGGTCTATGAGCTGGAGCGCGCTTCAAGCAGCAGCGGTACCTGTGCGCCCCCG 186
|||||
Qy 4975 aacgcgaccagctgccagcgtgtgaaactcacctccacgcaggtcgaagatctggtcc 5034
|||||
Db 187 AACGCGACAGCTGGCCAGCGTGTCTGAACCTCACGTCACGAGGTCAAGATCTGGTTCC 246
|||||
Qy 5035 agaaccggcgctacaagtgcgaagcgagcgagcagcagaccagactctggagctggtggcg 5094
|||||
Db 247 AGAACCGCGCTACAAAGTGAAGCGGACGCGCAGGACCGACGACTCTGGAGCTGGTGGGC 306
|||||
Qy 5095 tgcctcccgccgcgcgcgcctgcgcgagatcgcggtgacgtgtcgtggtgcgcgagtcg 5154
|||||
Db 307 TGCCCCCGCGCGCGCGCGCTGCGCGCAGGATCGCGTGCAGTGTCTGTGCGCGGATG 366
|||||
Qy 5155 gcaagcctgctagggagcactcgccctacgcgctcctacgcgctggcggtggcctcaatc 5214
|||||
Db 367 GCAAGCCATGCTAGGGGACTCGCGCGCTACGCGCCTGCCCTACGCGGTGGGCTCAATC 426
|||||
Qy 5215 cctac 5219
|||||
Db 427 CCTAC 431

RESULT 6
AW001138
LOCUS
DEFINITION
  AW001138 542 bp mRNA linear EST 25-OCT-2000
  HEMBA1 Homo sapiens cDNA clone HEMBA1007016 3' , mRNA
  sequence.
ACCESSION
  AW001138
VERSION
  AW001138.1 GI:11007681
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 542)
  Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
  Yanamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
  , S. and Isogai,T.).
```


LOCUS AI337523 418 bp mRNA linear EST 13-FEB-1999
DEFINITION qq02a02.x1 Soares fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:1931306 3', mRNA sequence.
ACCESSION AI337523
VERSION AI337523.1 GI:4074450
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 606 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 366.
FEATURES
source
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1931306"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p7T73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACATCTGAAGTGGAGCGCGCAATTTTCTTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."
96 a 120 c 131 g 71 t

Query Match 5.9%; Score 399; DB 9; Length 418;
Best Local Similarity 100.0%; Pred. No. 8.9e-69;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5512 ggggagaagggtcccaacatgacctgagtcctcctgagtttgcattcactcctgcgg 5571
|||||
Db 400 GGGGAGAAGGGGCTCCCAACATGACCTTCTAGTCCCTCGATTGTCATTCTCTCGGG 341
|||||
Qy 5572 agacctaggaacttttctgtccacgcgcgtttgttcttcgcacggagagtttgtgg 5631
|||||
Db 340 AGACCTAGAACTTTTCTGTCCACGCGCTTTCTTCTTGGCAGCGGAGAGTTGTGG 281
|||||
Qy 5632 cggcgattatcagcgtgcaatgagtatcctgcagctggtgtcttagctgtccccc 5691
|||||
Db 280 CGGCGATTATCAGCGTGCAATGATGATCTGTGACCTGTGTCTTATGCTGCCCCCA 221
|||||
Qy 5692 ggagtgcctccgagatccatgggaccccccggttggaactgggaactggagctcgggac 5751
|||||
Db 220 GGAGTGCCCTCGGAGAGTCCATGGGACCCCCGGTTGGAACCTGGAGCTGAGCTCGGGAC 161
|||||
Qy 5752 gcagggcctgagatgtggcccccattccgcgagccagggccggcgccggtttgc 5811
|||||
Db 160 GCAGGGCCTGAGATGTGGCCGCCCATTCGCGAGCAGGCGCGCGCGCGCTTTGC 101
|||||
Qy 5812 tcttcgcgctgcgcgcgcgcgcaccccgatttatgttttacctatttctgttaa 5871
|||||
Db 100 TATCTGCCGCTGCCGCCGCCACGCCACCCACCCGATTTATGTTTTTACCTATTGCTGTAA 41
|||||

Qy 5872 gaaatgacgtcccccttccattaaagagagtcgttga 5910
|||||
Db 40 GAAATGACGATCCCTTCCCATTAAGAGAGTGGTTGA 2
|||||

RESULT 9
BI834117 806 bp mRNA linear EST 04-OCT-2001
LOCUS 603085284F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5224365 5',
DEFINITION mRNA sequence.
ACCESSION BI834117
VERSION BI834117.1 GI:15945667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 806)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11564 row: g column: 22
High quality sequence stop: 736.
FEATURES
source
1..806
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5224365"
/clone_lib="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."
135 a 290 c 272 g 109 t

BASE COUNT 135 a 290 c 272 g 109 t
ORIGIN

Query Match 5.8%; Score 393; DB 10; Length 806;
Best Local Similarity 100.0%; Pred. No. 6e-68;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2895 gccgtgggcagcgccgtttctgcgcacacctggcgctgtgagactggcgctgcaccca 2954
|||||
Db 53 GCCGTGGGCGAGCGCGCTTCTGCCGCCACCTGGCGCTGTGAGACTGGCGTGCACCA 112
|||||
Qy 2955 tgttccccagccctgtctcacgcccacgcccttctcagtcacaaagacatcctaaacctgg 3014
|||||
Db 113 TGTTCGCCAGCCCTGTCTCAGCGCCACGCCCTTCTCAGTCAAGACATCTCTAAACCTGG 172
|||||
Qy 3015 aacagcagcagcgagcctgctgcgcgcagagactctctcccgccctggagggcagccc 3074
|||||
Db 173 AACAGCAGCGCAGCGCTGGCTGCCGCCGAGAGCTCTCTGCCGCCCTGGAGCGACCC 232
|||||
Qy 3075 tggegccctccctcctgcacgtgcgcgccttcaagccagagcgctacgctggcgccgag 3134
|||||
Db 233 TGGCGCCCTCCTCCTGCTGCTGCGCGCTTCAAGCCAGAGGCGCTACGCTGGGCCCGAGG 292
|||||


```
RESULT 15
BG115100
LOCUS
DEFINITION BG115100 938 bp mRNA linear EST 30-JAN-2001
602315938F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416512 5',
mRNA sequence.
ACCESSION BG115100
VERSION BG115100.1 GI:12608606
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 938)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10147 row: k column: 09
High quality sequence stop: 722.
FEATURES
Location/Qualifiers
source 1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4416512"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: Not1; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 189 a 281 c 311 g 157 t
ORIGIN
```

Query Match 4.0%; Score 270; DB 10; Length 938;
Best Local Similarity 100.0%; Pred. No. 3.9e-44;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4994 cgtgctgaactcagtcacagcaggtcaagatctgttccagaaacggcgctacaagtg 5053
|||||
Db 1 CGTGCTGAAACTCAGTCCACGACAGGTCAAGATCTGTTCCAGAACCGCGCTACAGTG 60
|||||

QY 5054 caagcgacgcgcagacagactctggagctggtgggctgcccgccgcgcgcgc 5113
|||||
Db 61 CAAGCGGCGAGCGGAGGACAGACTCTGGAGCTGCTGGGGGTGCCCGCGCGCGCGCC 120
|||||

QY 5114 gcctgccgcagaggtcgcgggtgcagtgctggtgcgagtggaagccatgcctagggga 5173
|||||
Db 121 GCCTGCCCGCAGGATCGCGGTGCCAGTCTGCTGCTGCCGATGGCAAGCCATGCCTAGGGGA 180
|||||

QY 5174 ctggcgccctacgcgctgctgctacgcggtgggctcaatccctacggtataacgccta 5233
|||||
Db 181 CTCGGCGCCCTACGCGCTGCTACGCGGTGGGCTCAATCCCTACGGTTATAACGCCTA 240
|||||

QY 5234 cccgcctatccgggttacgcgcgcgcgc 5263
|||||
Db 241 CCCGCGCTATCCGGTTACGCGCGCGCGC 270
|||||


```
* 54524 54623: gap of unknown length
* 54624 68414: contig of 13791 bp in length
* 68415 68514: gap of unknown length
* 68515 72850: contig of 4336 bp in length
* 72851 88213: gap of unknown length
* 72951 88213: contig of 15263 bp in length
* 88214 88314 98538: gap of unknown length
* 88314 98538: contig of 10225 bp in length
* 98539 98639 107774: gap of unknown length
* 98639 107774: contig of 9136 bp in length
* 107775 115145: gap of unknown length
* 107875 115145: contig of 7271 bp in length
* 115146 118552: gap of unknown length
* 115246 118552: contig of 3307 bp in length
* 118553 118653: gap of unknown length
* 118653 128198: contig of 9546 bp in length
* 128199 128298: gap of unknown length
* 128299 130123: contig of 1825 bp in length
* 130124 130223: gap of unknown length
* 130224 134038: contig of 3815 bp in length
* 134039 134138: gap of unknown length
* 134139 136470: contig of 2332 bp in length
* 136471 136570: gap of unknown length
* 136571 137545: contig of 975 bp in length.
FEATURES
    source
        1..137545
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone_lib="Caltech human BAC library C"
            /clone_id="CPC-281H14"
BASE COUNT 37277 a 32308 c 31466 g 34494 t 2000 others
ORIGIN

Query Match
Best Local Similarity 99.4%; Score 475; DB 2; Length 137545;
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agagaaatcattaccacattcacaaagacatagagagtgaacagtcactgtattgtt 60
    |||||
Db 36527 AGAGAAATCATTACCAGGATTCACAAAGAGCATAGAGAGTGTAAACAGTCACGTGATCTTGT 36586

QY 61 caaataggagaggttttttttcccttcctttttgtacacactgaccacaggaactgacag 120
    |||||
Db 36587 CAAATAGGGAGAGTGTGTTCCTTCCTTCCTTTTGTAAACACTGACCCACAGGACTGACAG 36646

QY 121 ttctaggaagcccttaccgaaataggaataaaatccttgcacctgtattgcaag 180
    |||||
Db 36647 TTCTAGGAAGCCCTTTACCGGAAATAGGAAATAAATCTCTTGCCACCTTGATTTGCAG 36706

QY 181 ggcaatgctaatttttttcttccagagctctcaaaaaaaaaaaaaaaaaaaccttac 240
    |||||
Db 36707 GGCAATGCTAATTTTCTTCTCCAGAGCTCTCAAAAAAAAAAAAAAAAAACCTTAC 36766

QY 241 taaaaacagggatcccgagtagcctcgatgtcccccaataaacgtaattatccaggc 300
    |||||
Db 36767 TAAAAACAGGGATCCCGATGTAGCCCTCGATGTCCCCCATTAACGGTAATATTTCAGGC 36826

QY 301 gtccgctcacactaatcttcaactgtcatcgagccgctgcccagcagatcactt 360
    |||||
Db 36827 GTCCGCTCACACTAATCTTTCAAACTGTCTATCGGAGCCGCTGGCCAGCAGATTCACCT 36886

QY 361 aacagcgtcccgagaccctcgttcccgagctctttcagcgagacatttaattgaatcgg 420
    |||||
Db 36887 AACAGCGCTCCAGACCCCTCGTTCCGAGCTCTTTTTCAGCGAGACATTAATTTGAATCGG 36946

QY 421 atgtggctcgttttgcagagatcccgccctcggcgtaggcattcctctccaacga 475
    |||||
Db 36947 ATGTGGCTCGTTTTCAGAGCTCACCGCCTCGGCCTAGGATAGGATCTCTCTCCAACGA 37001

RESULT 2
```

```
AL669924/c
LOCUS
DEFINITION Homo sapiens chromosome 6 clone Xxbac-250J14, *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces.
ACCESSION AL669924
VERSION AL669924.4 GI:18375903
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (sites)
REFERENCE
    Peck, A.
    Direct Submission
    Submitted (17-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
    Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
    humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
    On Jan 25, 2002 this sequence version replaced gi:18250865.
    ----- Genome Center
    Center: Wellcome Trust Sanger Institute
    Center code: SC
    Web site: http://www.sanger.ac.uk
    Contact: humquery@sanger.ac.uk
    ----- Project Information
    Center project name: bpg250J14
    ----- Summary Statistics
    Assembly program: XGAP4; version 4.5
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Consensus quality: 134624 bases at least Q40
    Consensus quality: 135417 bases at least Q30
    Consensus quality: 135772 bases at least Q20
    Insert size: 136057; sum-of-contigs
    Insert size: 115180; 6.1% error; agarose-fp
    Quality coverage: 7.81x in Q20 bases; sum-of-contigs Quality
    coverage: 9.23x in Q20 bases; agarose-fp
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 7 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    * 1 7523: contig of 7523 bp in length
    * 7524 7623: gap of 100 bp
    * 7624 41360: contig of 33737 bp in length
    * 41361 41460: gap of 100 bp
    * 41461 47515: contig of 6055 bp in length
    * 47516 47615: gap of 100 bp
    * 47616 52829: contig of 5214 bp in length
    * 52830 52929: gap of 100 bp
    * 52930 90493: contig of 37564 bp in length
    * 90494 90593: gap of 100 bp
    * 90594 109805: contig of 19212 bp in length
    * 109806 109905: gap of 100 bp
    * 109906 136657: contig of 26752 bp in length.
    Location/Qualifiers
        1..136657
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone_lib="Xxbac-250J14"
            /clone_id="Xxbac-250J14"
        1..7523
            /note="assembly fragment:01480
            fragment_chain:1
            clone_end:T7
            vector_side:left"
            7624..41360
                /note="assembly fragment:01718
                fragment_chain:1"
                41461..47515
                    misc_feature
                        source
                            1..136657
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /chromosome="6"
                                /clone_lib="Xxbac-250J14"
                                /clone_id="Xxbac-250J14"
                            1..7523
                                /note="assembly fragment:01480
                                fragment_chain:1
                                clone_end:T7
                                vector_side:left"
                                7624..41360
                                    /note="assembly fragment:01718
                                    fragment_chain:1"
                                    41461..47515
                                        misc_feature
```


Qy 121 ttctaggaagcccccttaccgaaataggaaataaatccttgcacaccttgatttgcaag 180
|||||
Db 11214 TTCTAGGAAGCCCCCTTACC CGAAATAGGAAATAAATCCTTGCACCTTGATTGCAAG 11273
|||||
Qy 181 ggcgaatgctaatttttttcttctccagagctctcaaaaaaaaaaaaaaaaaaccccttac 240
|||||
Db 11274 GGCAATGCTAATTTTCTTCTCCAGAGCTCTC - AAAAAAAAAAAAAAAAAACCTTAC 11332
|||||
Qy 241 taaaacagggatcccgatgtagcctcgtatgtcccccattaaacggtaatatatttcaggc 300
|||||
Db 11333 TAAAAACAGGGATCCCGAGTAGCCTCGATGTCCCCCATTAACGGTAATATTTCAGGC 11392
|||||
Qy 301 gtcgctcacactaatctttcaaaactgtcatcgcgagccgctgcccagcagattcaactt 360
|||||
Db 11393 GTCGCTCACACTAATCTTTCAAACGTCTATCGCGAGCGCCTGGCCAGCAGATTCACCTT 11452
|||||
Qy 361 aacagcgtcccgagacccctcgcttccagagctcttttccagcagacatttaattgaatcgg 420
|||||
Db 11453 AACAGCGTCCCGAGAGCCCTCGTTCCGAGCTCTTTTCAGCGAGACATTTAATTGAATCGG 11512
|||||
Qy 421 atgtggctgctttgcagagctcaaccgctcggtataggtatcctctctcaacgacac 478
|||||
Db 11513 ATGTGGCTCGTTGGCCAGCGTCACCGCCTCGGCGATAGGCATCCTCTCCAACGACAC 11570
|||||

Search completed: July 5, 2002, 10:27:10
Job time: 6432 sec

QY 361 aacagcgctccagagaccctcgcttcagagctcttttccagcgagacatttaattgaatcgg 420
|||||
Db 361 aacagcgctccagagaccctcgcttcagagctcttttccagcgagacatttaattgaatcgg 420
|||||
QY 421 atgtggctcttgcagagctcaccgctcgcgagataggcatcctctccaagacac 478
|||||
Db 421 atgtggctcttgcagagctcaccgctcgcgagataggcatcctctccaagacac 478
|||||

RESULT 2
AAS09962
ID AAS09962 standard; DNA; 6751 BP.
AC AAS09962;
DT 24-OCT-2001 (first entry)
XX Genomic DNA #2 encoding human Csx/Nkx2.5.
DE
XX Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
KW therapeutic; heart tissue; gene therapy; human; ds.
KW
XX Homo sapiens.
OS
XX WO200151006-A2.
PN
XX 19-JUL-2001.
PD
XX 16-JAN-2001; 2001WO-US01511.
PF
XX 14-JAN-2000; 2000US-0176419.
PR
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA
XX Lee IW, Izumo S;
PI
XX WPI; 2001-451809/48.
DR
XX
XX New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
XX Claim 29; Fig 4B; 66pp; English.
XX
XX The sequence represents the genomic sequence #2 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
XX Sequence 6751 BP; 1459 A; 2008 C; 1753 G; 1490 T; 41 other;

Query Match 100.0%; Score 478; DB 22; Length 6751;
Best Local Similarity 100.0%; Pred. No. 1.1e-175;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agagaaatcattaccgattcacaaaagacatagagtgtaacagtcactgatctgtt 60
|||||
Db 201 agagaaatcattaccgattcacaaaagacatagagtgtaacagtcactgatctgtt 260
|||||
QY 61 caaatagggagagtttttttccctcttcttttgaacacctgacccacagagactgacag 120
|||||
Db 261 caaatagggagagtttttttccctcttcttttgaacacctgacccacagagactgacag 320
|||||
QY 121 ttctaggaagcccttaccgaaaataggaataaaatccttgccacctgtatttgaag 180
|||||
Db 321 ttctaggaagcccttaccgaaaataggaataaaatccttgccacctgtatttgaag 380
|||||
QY 181 ggcaatgctaatttttttccagagctctcaaaaaaaaaaaaaaaaaacattac 240
|||||

Db 381 ggcaatgctaatttttttccagagctctcaaaaaaaaaaaaaaaaaacattac 440
QY 241 taaaaacagggatcccgatgtagcctgcgtcccccattaaacggttaattttcaggc 300
|||||
Db 441 taaaaacagggatcccgatgtagcctgcgtcccccattaaacggttaattttcaggc 500
|||||
QY 301 gtcgctcacactaatattttcaaaactgtcatcgagcgccgctggccagcagattcactt 360
|||||
Db 501 gtcgctcacactaatattttcaaaactgtcatcgagcgccgctggccagcagattcactt 560
|||||
QY 361 aacagcgctccagagaccctcgcttcagagctcttttccagcgagacatttaattgaatcgg 420
|||||
Db 561 aacagcgctccagagaccctcgcttcagagctcttttccagcgagacatttaattgaatcgg 620
|||||
QY 421 atgtggctcttgcagagctcaccgctcgcgagataggcatcctctccaagacac 478
|||||
Db 621 atgtggctcttgcagagctcaccgctcgcgagataggcatcctctccaagacac 678
|||||

Search completed: July 5, 2002, 10:31:08
Job time: 3910 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

JM sequence - nucleic search, using sw model

Run on: July 5, 2002, 08:41:18 : Search time 44.79 seconds
(without alignments)
2621.407 Million cell updates/sec

Title: US-09-761-466-6
Perfect score: 478
Sequence: : aqaqaatcaatcaacaaat.....agatcctctctcaaacagaa 478

Scoring table: 119..Nm*
Gapop 60,0 , Gapext 60,0

Searched: 88533 seqs, 12281652 residues

Word size : 50

Total number of hits satisfying chosen parameters: 0

Minimum db seq length: 0
Maximum db seq length: 200000000

Post-processing: Listing first 45 summaries

Database : ISSUED_Patents_NA:*
1: /cqn2_6/ptodata/2/ina/5A_0.MB.seq:*
2: /cqn2_6/ptodata/2/ina/5B_0.MB.seq:*
3: /cqn2_6/ptodata/2/ina/5A_0.MB.seq:*
4: /cqn2_6/ptodata/2/ina/5B_0.MB.seq:*
5: /cqn2_6/ptodata/2/ina/PTUS_0.MB.seq:*
6: /cqn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
.....					

No matches found

Search completed: July 5, 2002, 10:27:15
Job time: 6457 sec

GeneCore version 4.5
Copyright (c) 1993 - 2000 Compugen, Ltd.

EM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 08:05:28 : Search time 1616.98 seconds
(without alignments)
3489.873 Million cell updates/sec

Title: US-09-761-466-6

Period score: 478

Sequence: 1 aqaaataatcattaccgaatt.....agcaatcctctcvaactaacac 478

Scoring table: GLLIGG_NRC

Gapop 60.0 , Gapext 60.0

Searched: 13756207 seqs, 6748477642 residues

Word size: 5

Total number of hits satisfying chosen parameters: 0

Minimum hit seq length: 0

Maximum hit seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: EST:
1: em_ostba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estr:*
9: ab_est1:*
10: ab_est2:*
11: ab_hic:*
12: ab_488:*
13: em_488_hum:*
14: em_488_inz:*
15: em_488_pln:*
16: em_488_vrt:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB	ID	Description
---------------	----------------	-------	--------	----	----	-------------

No matches found

Search completed: July 5, 2002, 09:56:05
Job time: 0577 sec

GenCore version 4.5
copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 06:24:08 ; Search time 1776.56 Seconds
(without alignments)
5610.479 Million cell updates/sec

Title: US-09-761-466-6
Perfect score: 478
Sequence: 1 agagaaatcattaccggatt.....ggacatctctccaaacacac 478

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1777656 seqs, 10463264293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum iB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba:*
- 2: gb_hq:*
- 3: gb_in:*
- 4: gb_ma:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pt:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_ma:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_un:*
- 28: em_vl:*
- 29: em_hq_hum:*
- 30: em_hq_inv:*
- 31: em_hq_inv:*
- 32: em_hq_inv:*
- 33: em_hq_inv:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	476.4	95.7	137545	2	AC008412
2	466	97.5	136657	2	AL669924
3	466	97.5	144702	2	AL669924
4	159.8	41.8	8117	10	AF091451
5	159.8	41.8	219180	2	AC092251
6	182.2	38.1	8648	10	AF084133
7	44.2	9.2	191494	2	AL591489
8	41.8	8.7	77483	9	AC027171
9	41.8	8.7	141742	9	AC092521
10	41	8.5	168698	9	AC091448
11	40.8	8.5	184664	9	AC092424
12	40.8	8.5	195430	9	AC026447
13	40.4	8.5	6152	5	AF283102
14	40.2	8.4	48488	3	AF106589
15	40.2	8.4	200000	2	AL007700
16	40	8.4	182695	2	AC015706
17	40	8.4	202955	9	AL456478
18	39.6	8.3	535	3	AF279411
19	39.6	8.3	535	3	AF279421
20	39.6	8.3	549	3	AF097010
21	39.6	8.3	54770	9	HS4012
22	39.6	8.3	149656	2	AC015573
23	39.6	8.3	170392	2	AC021523
24	39.6	8.3	180719	2	AC087692
25	39.6	8.3	181231	2	AC091625
26	39.2	8.2	120766	9	AC004150
27	39.2	8.2	143422	2	AC086636
28	39.2	8.2	179671	2	AF002952
29	39	8.2	144542	2	AC027146
30	39	8.2	151533	8	AC090873
31	39	8.2	155555	2	AL591710
32	39	8.2	162066	2	AF002986
33	39	8.2	162591	9	AL451839
34	39	8.2	170682	2	AC024428
35	39	8.2	172905	2	AC025130
36	39	8.2	174639	2	AC067862
37	39	8.2	193902	2	AL591312
38	38.8	8.1	63310	2	AC087526
39	38.8	8.1	104547	9	AC008416
40	38.8	8.1	144744	2	AC015674
41	38.8	8.1	168135	9	AC025882
42	38.8	8.1	217532	2	AL445543
43	38.6	8.1	535	3	AF279416
44	38.6	8.1	82896	8	AF000420
45	38.6	8.1	164644	2	AC097448

ALIGNMENTS

RESULT	1
AC008412	
LOCUS	AC008412 Homo sapiens chromosome 5 clone c10-24P14, WIKING DRAFT SEQUENC
DEFINITION	21 ordered pieces.
ACCESSION	AC008412
VERSION	AC008412.5 GI:5255970
KEYWORDS	HIG: HIGS_PHASE2: HIGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 137545)
TITLE	Sequencing of Human Chromosome 5
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 137545)
AUTHORS	1 (bases 1 to 137545)
TITLE	Direct Submission
JOURNAL	Submitted (03-AUG-1999) Probert Gen Sequencing Facility, DOE Joint Genome Institute, 2600 Walnut Creek, CA 94596, USA
COMMENT	on Jul 18, 2000 this sequence version replaced gi:7708438. -----Genome Center

Insert size: 14657; sum-of-contigs
 quality: 14657; 6.1% error; adarose-tp
 quality: coverage: 7.81x in 920 bases; sum-of-contigs quality
 coverage: 9.24x in 920 bases; adarose-tp
 coverage: 9.24x in 920 bases; adarose-tp

NOTE: This is a 'working draft' sequence. It currently
 consists of 7 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence,
 as soon as it is available and the accession number will
 be preserved.

1 7524: contig of 7524 bp in length
 7524 7624: gap of 100 bp
 7624 41460: contig of 34737 bp in length
 41460 41461: gap of 100 bp
 41461 47515: contig of 6055 bp in length
 47515 47516: gap of 100 bp
 47516 52829: contig of 5214 bp in length
 52829 52930: gap of 100 bp
 52930 90594: contig of 37564 bp in length
 90594 90595: gap of 100 bp
 90595 109805: contig of 19212 bp in length
 109805 109806: gap of 100 bp
 109806 136657: contig of 26752 bp in length.

FEATURES
 source
 1..136657
 /organism "Homo sapiens"
 /db_xref:"taxon:9606"
 /chromosome:"6"
 /clone:"XXbac-25c014"
 /clone_lib:"CHORI-901"
 1..7524
 /note:"assembly fragment:01480"
 fragment_chain:1
 clone_end:17
 vector_side:"left"
 7524..41460
 /note:"assembly fragment:01718"
 fragment_chain:1
 41461..47515
 /note:"assembly fragment:00204"
 fragment_chain:1
 47516..52829
 /note:"assembly fragment:02002"
 fragment_chain:1
 52930..90594
 /note:"assembly fragment:00959"
 fragment_chain:1
 90595..109805
 /note:"assembly fragment:01522"
 fragment_chain:1
 109906..136657
 /note:"assembly fragment:01742"
 fragment_chain:1
 136658..42573
 42574..92173
 92174..92175
 92176..92177
 92178..92179
 92180..92181
 92182..92183
 92184..92185
 92186..92187
 92188..92189
 92190..92191
 92192..92193
 92194..92195
 92196..92197
 92198..92199
 92200..92201
 92202..92203
 92204..92205
 92206..92207
 92208..92209
 92210..92211
 92212..92213
 92214..92215
 92216..92217
 92218..92219
 92220..92221
 92222..92223
 92224..92225
 92226..92227
 92228..92229
 92230..92231
 92232..92233
 92234..92235
 92236..92237
 92238..92239
 92240..92241
 92242..92243
 92244..92245
 92246..92247
 92248..92249
 92250..92251
 92252..92253
 92254..92255
 92256..92257
 92258..92259
 92260..92261
 92262..92263
 92264..92265
 92266..92267
 92268..92269
 92270..92271
 92272..92273
 92274..92275
 92276..92277
 92278..92279
 92280..92281
 92282..92283
 92284..92285
 92286..92287
 92288..92289
 92290..92291
 92292..92293
 92294..92295
 92296..92297
 92298..92299
 92300..92301
 92302..92303
 92304..92305
 92306..92307
 92308..92309
 92310..92311
 92312..92313
 92314..92315
 92316..92317
 92318..92319
 92320..92321
 92322..92323
 92324..92325
 92326..92327
 92328..92329
 92330..92331
 92332..92333
 92334..92335
 92336..92337
 92338..92339
 92340..92341
 92342..92343
 92344..92345
 92346..92347
 92348..92349
 92350..92351
 92352..92353
 92354..92355
 92356..92357
 92358..92359
 92360..92361
 92362..92363
 92364..92365
 92366..92367
 92368..92369
 92370..92371
 92372..92373
 92374..92375
 92376..92377
 92378..92379
 92380..92381
 92382..92383
 92384..92385
 92386..92387
 92388..92389
 92390..92391
 92392..92393
 92394..92395
 92396..92397
 92398..92399
 92400..92401
 92402..92403
 92404..92405
 92406..92407
 92408..92409
 92410..92411
 92412..92413
 92414..92415
 92416..92417
 92418..92419
 92420..92421
 92422..92423
 92424..92425
 92426..92427
 92428..92429
 92430..92431
 92432..92433
 92434..92435
 92436..92437
 92438..92439
 92440..92441
 92442..92443
 92444..92445
 92446..92447
 92448..92449
 92450..92451
 92452..92453
 92454..92455
 92456..92457
 92458..92459
 92460..92461
 92462..92463
 92464..92465
 92466..92467
 92468..92469
 92470..92471
 92472..92473
 92474..92475
 92476..92477
 92478..92479
 92480..92481
 92482..92483
 92484..92485
 92486..92487
 92488..92489
 92490..92491
 92492..92493
 92494..92495
 92496..92497
 92498..92499
 92500..92501
 92502..92503
 92504..92505
 92506..92507
 92508..92509
 92510..92511
 92512..92513
 92514..92515
 92516..92517
 92518..92519
 92520..92521
 92522..92523
 92524..92525
 92526..92527
 92528..92529
 92530..92531
 92532..92533
 92534..92535
 92536..92537
 92538..92539
 92540..92541
 92542..92543
 92544..92545
 92546..92547
 92548..92549
 92550..92551
 92552..92553
 92554..92555
 92556..92557
 92558..92559
 92560..92561
 92562..92563
 92564..92565
 92566..92567
 92568..92569
 92570..92571
 92572..92573
 92574..92575
 92576..92577
 92578..92579
 92580..92581
 92582..92583
 92584..92585
 92586..92587
 92588..92589
 92590..92591
 92592..92593
 92594..92595
 92596..92597
 92598..92599
 92600..92601
 92602..92603
 92604..92605
 92606..92607
 92608..92609
 92610..92611
 92612..92613
 92614..92615
 92616..92617
 92618..92619
 92620..92621
 92622..92623
 92624..92625
 92626..92627
 92628..92629
 92630..92631
 92632..92633
 92634..92635
 92636..92637
 92638..92639
 92640..92641
 92642..92643
 92644..92645
 92646..92647
 92648..92649
 92650..92651
 92652..92653
 92654..92655
 92656..92657
 92658..92659
 92660..92661
 92662..92663
 92664..92665
 92666..92667
 92668..92669
 92670..92671
 92672..92673
 92674..92675
 92676..92677
 92678..92679
 92680..92681
 92682..92683
 92684..92685
 92686..92687
 92688..92689
 92690..92691
 92692..92693
 92694..92695
 92696..92697
 92698..92699
 92700..92701
 92702..92703
 92704..92705
 92706..92707
 92708..92709
 92710..92711
 92712..92713
 92714..92715
 92716..92717
 92718..92719
 92720..92721
 92722..92723
 92724..92725
 92726..92727
 92728..92729
 92730..92731
 92732..92733
 92734..92735
 92736..92737
 92738..92739
 92740..92741
 92742..92743
 92744..92745
 92746..92747
 92748..92749
 92750..92751
 92752..92753
 92754..92755
 92756..92757
 92758..92759
 92760..92761
 92762..92763
 92764..92765
 92766..92767
 92768..92769
 92770..92771
 92772..92773
 92774..92775
 92776..92777
 92778..92779
 92780..92781
 92782..92783
 92784..92785
 92786..92787
 92788..92789
 92790..92791
 92792..92793
 92794..92795
 92796..92797
 92798..92799
 92800..92801
 92802..92803
 92804..92805
 92806..92807
 92808..92809
 92810..92811
 92812..92813
 92814..92815
 92816..92817
 92818..92819
 92820..92821
 92822..92823
 92824..92825
 92826..92827
 92828..92829
 92830..92831
 92832..92833
 92834..92835
 92836..92837
 92838..92839
 92840..92841
 92842..92843
 92844..92845
 92846..92847
 92848..92849
 92850..92851
 92852..92853
 92854..92855
 92856..92857
 92858..92859
 92860..92861
 92862..92863
 92864..92865
 92866..92867
 92868..92869
 92870..92871
 92872..92873
 92874..92875
 92876..92877
 92878..92879
 92880..92881
 92882..92883
 92884..92885
 92886..92887
 92888..92889
 92890..92891
 92892..92893
 92894..92895
 92896..92897
 92898..92899
 92900..92901
 92902..92903
 92904..92905
 92906..92907
 92908..92909
 92910..92911
 92912..92913
 92914..92915
 92916..92917
 92918..92919
 92920..92921
 92922..92923
 92924..92925
 92926..92927
 92928..92929
 92930..92931
 92932..92933
 92934..92935
 92936..92937
 92938..92939
 92940..92941
 92942..92943
 92944..92945
 92946..92947
 92948..92949
 92950..92951
 92952..92953
 92954..92955
 92956..92957
 92958..92959
 92960..92961
 92962..92963
 92964..92965
 92966..92967
 92968..92969
 92970..92971
 92972..92973
 92974..92975
 92976..92977
 92978..92979
 92980..92981
 92982..92983
 92984..92985
 92986..92987
 92988..92989
 92990..92991
 92992..92993
 92994..92995
 92996..92997
 92998..92999
 93000..93001
 93002..93003
 93004..93005
 93006..93007
 93008..93009
 93010..93011
 93012..93013
 93014..93015
 93016..93017
 93018..93019
 93020..93021
 93022..93023
 93024..93025
 93026..93027
 93028..93029
 93030..93031
 93032..93033
 93034..93035
 93036..93037
 93038..93039
 93040..93041
 93042..93043
 93044..93045
 93046..93047
 93048..93049
 93050..93051
 93052..93053
 93054..93055
 93056..93057
 93058..93059
 93060..93061
 93062..93063
 93064..93065
 93066..93067
 93068..93069
 93070..93071
 93072..93073
 93074..93075
 93076..93077
 93078..93079
 93080..93081
 93082..93083
 93084..93085
 93086..93087
 93088..93089
 93090..93091
 93092..93093
 93094..93095
 93096..93097
 93098..93099
 93100..93101
 93102..93103
 93104..93105
 93106..93107
 93108..93109
 93110..93111
 93112..93113
 93114..93115
 93116..93117
 93118..93119
 93120..93121
 93122..93123
 93124..93125
 93126..93127
 93128..93129
 93130..93131
 93132..93133
 93134..93135
 93136..93137
 93138..93139
 93140..93141
 93142..93143
 93144..93145
 93146..93147
 93148..93149
 93150..93151
 93152..93153
 93154..93155
 93156..93157
 93158..93159
 93160..93161
 93162..93163
 93164..93165
 93166..93167
 93168..93169
 93170..93171
 93172..93173
 93174..93175
 93176..93177
 93178..93179
 93180..93181
 93182..93183
 93184..93185
 93186..93187
 93188..93189
 93190..93191
 93192..93193
 93194..93195
 93196..93197
 93198..93199
 93200..93201
 93202..93203
 93204..93205
 93206..93207
 93208..93209
 93210..93211
 93212..93213
 93214..93215
 93216..93217
 93218..93219
 93220..93221
 93222..93223
 93224..93225
 93226..93227
 93228..93229
 93230..93231
 93232..93233
 93234..93235
 93236..93237
 93238..93239
 93240..93241
 93242..93243
 93244..93245
 93246..93247
 93248..93249
 93250..93251
 93252..93253
 93254..93255
 93256..93257
 93258..93259
 93260..93261
 93262..93263
 93264..93265
 93266..93267
 93268..93269
 93270..93271
 93272..93273
 93274..93275
 93276..93277
 93278..93279
 93280..93281
 93282..93283
 93284..93285
 93286..93287
 93288..93289
 93290..93291
 93292..93293
 93294..93295
 93296..93297
 93298..93299
 93300..93301
 93302..93303
 93304..93305
 93306..93307
 93308..93309
 93310..93311
 93312..93313
 93314..93315
 93316..93317
 93318..93319
 93320..93321
 93322..93323
 93324..93325
 93326..93327
 93328..93329
 93330..93331
 93332..93333
 93334..93335
 93336..93337
 93338..93339
 93340..93341
 93342..93343
 93344..93345
 93346..93347
 93348..93349
 93350..93351
 93352..93353
 93354..93355
 93356..93357
 93358..93359
 93360..93361
 93362..93363
 93364..93365
 93366..93367
 93368..93369
 93370..93371
 93372..93373
 93374..93375
 93376..93377
 93378..93379
 93380..93381
 93382..93383
 93384..93385
 93386..93387
 93388..93389
 93390..93391
 93392..93393
 93394..93395
 93396..93397
 93398..93399
 93400..93401
 93402..93403
 93404..93405
 93406..93407
 93408..93409
 93410..93411
 93412..93413
 93414..93415
 93416..93417
 93418..93419
 93420..93421
 93422..93423
 93424..93425
 93426..93427
 93428..93429
 93430..93431
 93432..93433
 93434..93435
 93436..93437
 93438..93439
 93440..93441
 93442..93443
 93444..93445
 93446..93447
 93448..93449
 93450..93451
 93452..93453
 93454..93455
 93456..93457
 93458..93459
 93460..93461
 93462..93463
 93464..93465
 93466..93467
 93468..93469
 93470..93471
 93472..93473
 93474..93475
 93476..93477
 93478..93479
 93480..93481
 93482..93483
 93484..93485
 93486..93487
 93488..93489
 93490..93491
 93492..93493
 93494..93495
 93496..93497
 93498..93499
 93500..93501
 93502..93503
 93504..93505
 93506..93507
 93508..93509
 93510..93511
 93512..93513
 93514..93515
 93516..93517
 93518..93519
 93520..93521
 93522..93523
 93524..93525
 93526..93527
 93528..93529
 93530..93531
 93532..93533
 93534..93535
 93536..93537
 93538..93539
 93540..93541
 93542..93543
 93544..93545
 93546..93547
 93548..93549
 93550..93551
 93552..93553
 93554..93555
 93556..93557
 93558..93559
 93560..93561
 93562..93563
 93564..9356

* 116347 116446: gap of 100 bp
* 116447 131258: contig of 14812 bp in length
* 131259 131358: gap of 100 bp
* 131359 144702: contig of 13344 bp in length.

FEATURES
source

Location/Qualifiers
1. .144702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXbac-116A1"
/clone_lib="CHORI-501"
1. .75928
/note="assembly_fragment:03023
fragment_chain:1
clone_end:T7
vector_side:left"
76029. .116346
/note="assembly_fragment:03493
fragment_chain:1"
116447. .131258
/note="assembly_fragment:03524
fragment_chain:1"
131359. .144702
/note="assembly_fragment:01454
fragment_chain:1
clone_end:SP6
vector_side:right"
41517 a 33170 c 32886 g 36829 t 300 others

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 41517 a 33170 c 32886 g 36829 t 300 others
ORIGIN

Query Match 97.5%; Score 466; DB 2; Length 144702;
Best Local Similarity 99.8%; Pred. No. 8.3e-118;
Matches 477; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 agagaaatcattaccgattcacaaagacatagagagtgaacagtcactgtctgtt 60
Db 11094 AGAAAAATCATTACCGATTACAAAGAGCATAGAGAGTGAACAGTCACATCTTGT 11153
Qy 61 caaataggagagtttttttccctctttttgttaacacctgaccacagactgacag 120
Db 11154 CAAATAGGAGAGTGTTCCTTCCTCTTTTGTAAACACCTGACCCACAGACTGACAG 11213
Qy 121 ttctaggaagccccctaccgaaataggaataaaatcccttgccacctgattgcaag 180
Db 11214 TTCTAGGAAGCCCCCTTACCAGAAATAGGAATAAATCTTGCACCTTGATTGCAAG 11273
Qy 181 ggaatgctaatttttttcttccagagctctcaaaaaaataaccccttac 240
Db 11274 GGCAATGCTAATTTTCTTCTCCAGAGCTCTC-AAAAAATAAATAACCTTAC 11332
Qy 241 taaaaacagggatcccgatgtagctgtagtgcctgtagtccccattaaacgggtaatttcaggc 300
Db 11333 TAAAAACAGGGATCCGGATGATGCTGCTGATGTCCTCCCATTTAACGGTAAATTTTCAGCG 11392
Qy 301 gtccgctcacactaatctttcaaaactgtcatcgagcgccgtgagcagattcaactt 360
Db 11393 GTCCGCTCACACTAATCTTTCAAACTGTATCGGAGCGCCTGGCCAGCAGATTCACTT 11452
Qy 361 aacagcgctccagagaccctggtccagagctcttttcagcagacatttaataatggg 420
Db 11453 AACAGCGCTCCAGAGCCCTCGTTCGGAGCTCTTTTCAGCGAGACATTTAATTGAATCGG 11512
Qy 421 atgtggctcggttgcagacgtcacccgctcgggataggcatcctctccacgacac 478
Db 11513 ATGTGGCTCGTTTGGCAGACGTACCAGCTCGGCGATAGGCATCTCTCCAACGACAC 11570

RESULT 4

AF091351
LOCUS AF091351 8117 bp DNA linear ROD 23-DEC-1998
DEFINITION Mus musculus homeobox protein NKX-2.5 (Nkx-2.5) gene, complete cds.
ACCESSION AF091351

AF091351.1 GI:4050015
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 8117)
Searcy,R.D., Vincent,E.B., Liberatore,C.M. and Yutzey,K.E.
A GATA-dependent nkx-2.5 regulatory element activates early cardiac
gene expression in transgenic mice
Development 125 (22), 4461-4470 (1998)
98453405
9778505
2 (bases 1 to 8117)
Searcy,R.D. and Yutzey,K.E.
Direct Submission
Submitted (08-SEP-1998) Molecular Cardiovascular Biology.
Children's Hospital Research Foundation, 3333 Burnet Avenue,
Cincinnati, OH 45229, USA
Location/Qualifiers
1. .8117
/organism="Mus musculus"
/strain="129/J"
/db_xref="taxon:10090"
1. .505
/gene="Nkx-2.5"
/note="GATA-dependent"
1. .5992
/gene="Nkx-2.5"
join(3057. .3629,5005. .5992)
/gene="Nkx-2.5"
/product="homeobox protein NKX-2.5"
join(3299. .3629,5005. .5630)
/gene="Nkx-2.5"
/codon_start=1
/product="homeobox protein NKX-2.5"
/protein_id="AAC97934.1"
/db_xref="GI:4050016"
/translation="MFPSPALTPTFVSKDILNLEQOORSLASGDLSEATLAPAS
CMLAAFKPEAYSGPEAAASGLAEKRAEMGPAPSPKCPAPPAAPFPYAGYGDPPA
KDPADKELCALOKAVELDKAETGAERPRARRRRKRVLFSAOVYELERRFKQOR
YLSAPERDOLASVLKLTSTQVKINFORNYKCKRQODOTLELLGPPPPRIAPV
LVROGKCLGDPAPAFAYGLNAYGNAIYPTPSYGGAAACSPGYSCAAYPAAPAAQ
PPAASANSFNVFGVDLNTVQSPMPGNSGVSTLHGIRAW"
BASE COUNT 1831 a 2267 c 2109 g 1909 t 1 others
ORIGIN
Query Match 41.8%; Score 199.8; DB 10; Length 8117;
Best Local Similarity 74.0%; Pred. No. 1.9e-44;
Matches 328; Conservative 0; Mismatches 92; Indels 23; Gaps 5;
Qy 36 gagtgtaacagtcactgatctgttcaaataggagaggttttttccctcccttttgt 95
Db 885 GGGTGGGAAGTCATGATTTTGTTCAAAATTAGAGAGTTT-CTTCTTTTCIT 939
Qy 96 aacacctgaccacacaggactgacagttcttaggaagcccccttaccocgaataaggaata 155
Db 940 GACACCTGACCACACAGATTTGTCAACTCT-GGAAGCCCTTATATCGGAAAA-AAGTGATA 997
Qy 156 aatccttgccactgatttgcaagggaatgctaattttttcttctccagagcttc 215
Db 998 AATCTTCACCATGATTTTGAAAAGGAACG-----TCCCTCTCAAACTATT 1047
Qy 216 aaaaaaataaaaaaaccttactaaaaacaggagatcccgatgtagcctgatgctc 275
Db 1048 TTTTAAAAAGACTTTAAACACGGATCCCGGATCGGCCCCCAATATAGTC-CC 1101
Qy 276 cccattaaacggtaattttcaggcgctccgctcacactaatctttcaaacgtcatcgcg 335
Db 1102 CCAATTAACGGTAATATTTTCAGGCTCAGCTCACACTAATCTTTTCAAACTGTTCATCGCG 1161
Qy 336 agccgcctggccagcagattacttaaacagcgtcccccagaccctcggttcccgagcttt 395


```

/note="alternatively spliced"
/product="cardiac homeobox transcription factor"
join(5100...5754,7132...8114)
/gene="Nkx2-5"
/note="alternatively spliced"
/product="cardiac homeobox transcription factor"
join(5424...5754,7132...7757)
/gene="Nkx2-5"
/function="involved in heart development"
/note="tinman homolog; Nkx family"
/codon_start=1
/product="cardiac homeobox transcription factor"
/protein_id="AAG38875.1"
/db_xref="GI:11602838"
/translation="MFPSPALTPFVSVDILNLEQQOQSLASGLSARLEATLAPAS
KMLAAKPEAYSGPEAAELAEAMGPAPSPKPCSPFAAPAFYPCGAYGDDPA
KDPADKKELCALQKAVELDKATDCAERPRARRRKPVLFSQAQVYELERFKQPR
YLSPAERDQLASVLKLTSTQVKLWFQNNRYKCKRQDQDTLELGGPPPPPRRIAYPV
LVRDQGLCDGPAAYAGPVAGVGNLAYGNAYPYPSYGGACSPGYSCAAYPAAAPPAH
APAAANSINVFNGVDLNTVQSPGMPQGNSTYLHGTRAW"
2041 a 2336 c 2179 g 2091 t 1 others

```

[illegible]

RESULT	7
AL591488	
LOCUS	191494 bp DNA linear HTG 22-NOV-2001
DEFINITION	Mus musculus chromosome 2 clone RP23-36p22, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AL591488
VERSION	AL591488.7 GI:17065727
KEYWORDS	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS      1 (sites)
TITLE        Pearce,A.
JOURNAL      Direct Submission
COMMENT      Submitted (21-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Nov 25, 2001 this sequence version replaced gi:17043820.
             ----- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ----- Project Information
             Center project name: BM36P22
             ----- Summary Statistics
             Sequencing program: XGAP4; version 4.5
             Sequencing vector: plasmid; L08752; 100% of reads
             Chemistry: Dye-terminator Big Dye; 99% of reads
             Chemistry: Dye-primer Big Dye; 0% of reads
             Consensus quality: 191473 bases at least Q40
             Consensus quality: 191494 bases at least Q30
             Consensus quality: 191494 bases at least Q20
             Insert size: 191494; sum-of-contigs
             Insert size: 168060; 8.7% error; agarose-fp
             Quality coverage: 16.18x in Q20 bases; sum-of-contigs Quality
             coverage: 18.44x in Q20 bases; agarose-fp
             -----
             * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.

FEATURES
source       Location/Qualifiers
             1..191494
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="2"
                /clone="RP23-36P22"
                /clone_lib="RPC1-23"
             1..191494
                /note="assembly_fragment:03311"

BASE COUNT   49688 a 46072 c 46038 g 49696 t

misc_feature

Query Match          9.2%; Score 44.2; DB 2; Length 191494;
Best Local Similarity 56.6%; Pred. No. 0.19;
Matches 82, Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 154 taatccttgccaccttgatttgcgaaggcaatgctaatttttttttcctccagagctc 213
    ||| |||| |||||| ||||||| ||||||| |||| | ||| |||
Db 159235 TACAAGCTTGAGATCTTGTATTCCAAAGGGCAATTCCTTTGCTTTCTTTTTCGCCAAGGTT 159294
    ||| ||||||| ||||||| ||||||| ||||||| |||| | ||| |||

Qy 214 tcacaaaaaaacccccctactataaacacgggatcccgcgaagtaccctcgatgt 273
    ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||| | ||| |||
Db 159295 ACATTAAAAAAAATAAACCTACACTAACCCCGATGGGTCTTAAGTCAAACC 159354
    ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||| | ||| |||

Qy 274 ccccatataacggtaatatttcag 298
    ||| ||| ||||| ||||| |||

Db 159355 CAGCCTTGCCACGGTACTCTCCGAG 159379

RESULT 8
AC027171/c
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-404F23 map 2, LOW-PASS
SEQUENCE SAMPLING.
AC027171
AC027171.1 GI:7331490
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE

1 (bases 1 to 77483)
Arrhen,B., Linton,L., Nustbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boudjakly,L., Boudhalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Gallimore,A., Cooke,P., deArellano,K., Dewar,K., Diaz,J.S.,
Lodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gade,D.,
Galaun,I., Gardyna,S., Glode,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karakas,A.,
Klein,J., Lakouque,K., Lamazares,R., Landers,I., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McGrath,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye-S., Theodore,J., Tirrell,A., Travers,M., Trifillio,J.,
Vassiliev,H., Viel,R., V.D.A., Wilson,B., Wu,X., Wyman,D., Ye,W.Z.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE

Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:

<http://ftp.genome.washington.edu/M/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRK

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6875

Center clone name: 404_F_23

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 784 883: contig of 783 bp in length
* 884 1715: contig of 842 bp in length
* 1716 1815: gap of 100 bp
* 1816 2635: contig of 820 bp in length
* 2636 2735: gap of 100 bp
* 2736 3523: contig of 788 bp in length
* 3524 3623: gap of 100 bp
* 3624 4447: contig of 814 bp in length
* 4438 4537: gap of 100 bp
* 4538 5364: contig of 826 bp in length
* 5364 5453: gap of 100 bp
* 5464 6265: contig of 802 bp in length
* 6266 6365: gap of 100 bp
* 6366 7189: contig of 824 bp in length
* 7190 7289: gap of 100 bp
* 7290 8104: contig of 815 bp in length
* 8105 8204: gap of 100 bp
* 8205 9008: contig of 804 bp in length
* 9009 9108: gap of 100 bp
* 9109 9939: contig of 831 bp in length

* 9940 10039: gap of 100 bp
* 10040 10873: contig of 834 bp in length
* 10874 10973: gap of 100 bp
* 10974 11801: contig of 828 bp in length
* 11802 11901: gap of 100 bp
* 11902 12697: contig of 796 bp in length
* 12698 12797: gap of 100 bp
* 12798 13591: contig of 794 bp in length
* 13592 14391: gap of 100 bp
* 14392 14595: contig of 814 bp in length
* 14596 14695: gap of 100 bp
* 14696 15423: contig of 827 bp in length
* 15433 15533: gap of 100 bp
* 15534 16360: contig of 828 bp in length
* 16361 16460: gap of 100 bp
* 16461 17278: contig of 818 bp in length
* 17279 17378: gap of 100 bp
* 17379 18214: contig of 836 bp in length
* 18215 18314: gap of 100 bp
* 18315 19152: contig of 838 bp in length
* 19153 19252: gap of 100 bp
* 19253 20088: contig of 836 bp in length
* 20089 20188: gap of 100 bp
* 20189 20963: contig of 775 bp in length
* 20964 21063: gap of 100 bp
* 21064 21894: contig of 831 bp in length
* 21895 21994: gap of 100 bp
* 21995 22769: contig of 776 bp in length
* 22760 22859: gap of 100 bp
* 22860 23645: contig of 786 bp in length
* 23646 23745: gap of 100 bp
* 23746 24541: contig of 796 bp in length
* 24542 24641: gap of 100 bp
* 24642 25447: contig of 800 bp in length
* 25448 25547: gap of 100 bp
* 25548 26373: contig of 826 bp in length
* 26374 26473: gap of 100 bp
* 26474 27287: contig of 814 bp in length
* 27288 27387: gap of 100 bp
* 27388 28210: contig of 824 bp in length
* 28211 28310: gap of 100 bp
* 28311 29107: contig of 797 bp in length
* 29108 29207: gap of 100 bp
* 29208 30038: contig of 831 bp in length
* 30039 30138: gap of 100 bp
* 30139 30966: contig of 824 bp in length
* 30967 31066: gap of 100 bp
* 31067 31872: contig of 806 bp in length
* 31873 31972: gap of 100 bp
* 31973 32772: contig of 800 bp in length
* 32773 32872: gap of 100 bp
* 32874 33672: contig of 800 bp in length
* 33673 34587: contig of 916 bp in length
* 34588 34687: gap of 100 bp
* 34688 35506: contig of 814 bp in length
* 35507 35606: gap of 100 bp
* 35607 36418: contig of 812 bp in length
* 36419 36518: gap of 100 bp
* 36519 37340: contig of 822 bp in length
* 37341 37440: gap of 100 bp
* 37441 38266: contig of 826 bp in length
* 38267 38366: gap of 100 bp
* 38367 39191: contig of 825 bp in length
* 39192 39291: gap of 100 bp
* 39292 40075: contig of 784 bp in length
* 40076 40175: gap of 100 bp
* 40176 41007: contig of 830 bp in length
* 41008 41105: gap of 100 bp
* 41106 41919: contig of 814 bp in length
* 41920 42019: gap of 100 bp
* 42020 42822: contig of 803 bp in length
* 42823 43522: gap of 100 bp

AC026347
VERSION AC026347.17 GI:15920049
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195430)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louleghed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,N., Pickens,R., Pridmore,I., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 195430)
Worley,K.C.
Direct Submission
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 195430)
Worley,K.C.
Direct Submission
Submitted (31-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 4, 2001 this sequence version replaced gi:15809081.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	source	Location/Qualifiers
		1..195430
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="3"
		/clone="RP11-362A9"
repeat_region		500..521
		/rpt_family="AT-rich"
repeat_region		1223..1252
		/rpt_family="AT-rich"
repeat_region		3700..3720
		/rpt_family="AT-rich"
repeat_region		complement(4526..4693)
		/rpt_family="MLT13"
repeat_region		4805..4857
		/rpt_family="T-rich"
repeat_region		complement(5206..5501)
		/rpt_family="AluX"
repeat_region		complement(5690..5930)
		/rpt_family="MIR"
repeat_region		7018..7041
		/rpt_family="TTTTGn"
repeat_region		complement(8516..8653)
		/rpt_family="MIR"
repeat_region		8909..8937
		/rpt_family="AT-rich"
repeat_region		complement(9215..9519)
		/rpt_family="AluSg"
repeat_region		9789..9823
		/rpt_family="AT-rich"
repeat_region		10126..10425
		/rpt_family="AluYa5"
repeat_region		complement(11240..11535)
		/rpt_family="AluSg"
repeat_region		11572..11790
		/rpt_family="MIR"
repeat_region		complement(12590..12887)
		/rpt_family="AluSp"
repeat_region		14059..14080
		/rpt_family="(CA)n"
repeat_region		14081..14115
		/rpt_family="(TC)n"
repeat_region		15047..15081
		/rpt_family="(TTC)n"
repeat_region		15242..15269

KEYWORDS

SOURCE
ORGANISM

HTG.

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 48488)

The C. elegans Sequencing Consortium.

Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium

Science 282 (5396), 2012-2018 (1998)

99069613

2 (bases 1 to 48488)

Woesner,J., Graves,T. and Keppler,D.

The sequence of C. elegans cosmid Y44E3A

Unpublished

3 (bases 1 to 48488)

Waterston,R.

Direct Submission

Unpublished

4 (bases 1 to 48488)

Waterston,R.

Direct Submission

Submitted (15-NOV-1998) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

5 (bases 1 to 48488)

Waterston,R.

Direct Submission

Submitted (05-MAR-1999) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 48488)

Waterston,R.

Direct Submission

Submitted (30-JUN-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

7 (bases 1 to 48488)

Waterston,R.

Direct Submission

Submitted (03-JUL-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

8 (bases 1 to 48488)

Waterston,R.

Direct Submission

Submitted (16-NOV-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

On Jun 30, 2001 this sequence version replaced gi:3886084.

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

email: rwnematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program GeneFinder(p. Green and L. Hillier, ms in preparation).
Location/Qualifiers

source

1. 48488

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="Y44E3A"

complement(51..5319)

/gene="Y44E3A.2"

/note="ace-2"

complement(join(51..175,348..508,1022..1195,1721..2145,

3031..3267,3314..3542,3798..3973,4078..4281,5161..5319))

/gene="Y44E3A.2"

/note="C. elegans acetylcholinesterase (ACED-2)

(GB:AF025378); similar to carboxylesterases (Pfam:

PF00135, Score=363.2, E=4.9e-106, N=1); coded for by the

following C. elegans cDNAs: AF025378"

/codon_start=1

/product="Hypothetical protein Y44E3A.2"

/protein_id="AAC78228.2"

/db_xref="GI:14578252"

/translation="MRAPVIGRHLYHVFQFALVTLFIVRRIEPRISIVRGDHHVHTP

LGTRGVQTFDGAQVSAFLGVPAKPIGSRFRKMAEMIDRWGSELEARTLAKTCYL

TIDSAFPQFCAEMWNPQAISEDCLNMIWPDHDCSVMWLYGGFFSCTPSLDL

YGSVFAAKDHTIVVNYRLGPFGLFDGDDSPIQGNMGLMDQOLALRWVHENTIGAF

GGDRSVTLFGESAGSASTTAHLFAPNSHKYFRNLIAKSGSIISWASATPTMLDLS

FLAKKYNCSSPDMAIVKLSRVAHLVQAEADNISGIDGPPMTFAVVPVSSDANFF

QGDVFOKLANKQFKKDVNIIFGSVKDEGTYWLYMSLPKYFAFNHTISAEDPHNRA

LITRDHYEESMRAPMPYFAGSKVLVNAFMNSYEHVSTNVPFEYRDCVAFGLGDLFF

TCSLIDFADLLISNIFGVNYMYTYFYRSSANPWPQKMWGMVGHYELEYAFGOPYWRPHL

YDQTHLEDERKRLSIIIMQIWANFANTGRTDSFWQYKIERKAELELGTGTTLQGGKRRIL

SDVHGFGFRMIDEAKFAVQKNANDCTRTRKASTEDLTSSSTTYTYPISIIYLSILI

SYISL"

complement(15909..16429)

/gene="Y44E3A.3"

complement(join(15909..16140,16338..16429))

/gene="Y44E3A.3"

/note="contains similarity to thioredoxins (Pfam: PF00085,

Score=99.8, E=9.3e-29, N=1)"

/codon_start=1

/product="Hypothetical protein Y44E3A.3"

/protein_id="AAC78230.1"

/db_xref="GI:3886088"

/translation="MSIAIKDDGDEEFTFAEKKTPQVILFTFTASWCGPCQMIKPRVEE

LAAEHKDRLSILKIDVDEDCGDEEYINSMPTEFLIVDGIKKDQFSGANNTKFEEMV

KAAIQ"

complement(17707..21817)

/gene="Y44E3A.4"

complement(join(17707..17876,18825..19047,20000..20206,

20823..20984,21120..21160,21280..21397,21513..21677,

21779..21817))

/gene="Y44E3A.4"

/codon_start=1

/product="Hypothetical protein Y44E3A.4"

/protein_id="AAC78229.1"

/db_xref="GI:3886087"

/translation="MKDKIFKKIPTKKADPPPTIAPAAPSAPAPASQNLNPTSTPT

MPALAPSSALKSNFAALQKMSRKILQKNKNQFFLGLFLKFNKNSLKRKNEX

KPKNQKLSEKIDISIFSEFKNFFSKIETNKNCSFSENHSEKNOFSLFFQENFI

VLFLENMSNVLKITLNLKVVAPKVASIOKPDORSTSGSTTGEPSIDNSATDOVA

ELQHTKNRARPKNRPMVMNMRNSDESNGRLSSPTISITVNSMSASMFVPCA

KEISIPNSSATSPFFKSSSIAPPSASKPALRPIPSNLADDSSTSAIRKLASPD

SOKSPNLPSPDYSVRAEYDALLDRFRALESRLVAIEKIRF"

complement(24297..28465)

/gene="Y44E3A.5"

complement(join(24297..24584,25732..25935,26547..26890,

27694..27796,28448..28465))

/gene="Y44E3A.5"

/note="contains similarity to Src homology domain 3 (Pfam:

PF00018, Score=116.8, E=4e-31, N=2); coded for by the

following C. elegans cDNAs: yk341h3.5"

/codon_start=1

/product="Hypothetical protein Y44E3A.5"

/protein_id="AAC78231.1"

FEATURES

AC007760.1	GI:4996482	Human	HTG: HTGS PHASE1
1	1150	1150	1150
2	1151	1151	1151
3	1152	1152	1152
4	1153	1153	1153
5	1154	1154	1154
6	1155	1155	1155
7	1156	1156	1156
8	1157	1157	1157
9	1158	1158	1158
10	1159	1159	1159
11	1160	1160	1160
12	1161	1161	1161
13	1162	1162	1162
14	1163	1163	1163
15	1164	1164	1164
16	1165	1165	1165
17	1166	1166	1166
18	1167	1167	1167
19	1168	1168	1168
20	1169	1169	1169
21	1170	1170	1170
22	1171	1171	1171
23	1172	1172	1172
24	1173	1173	1173
25	1174	1174	1174
26	1175	1175	1175
27	1176	1176	1176
28	1177	1177	1177
29	1178	1178	1178
30	1179	1179	1179
31	1180	1180	1180
32	1181	1181	1181
33	1182	1182	1182
34	1183	1183	1183
35	1184	1184	1184
36	1185	1185	1185
37	1186	1186	1186
38	1187	1187	1187
39	1188	1188	1188
40	1189	1189	1189
41	1190	1190	1190
42	1191	1191	1191
43	1192	1192	1192
44	1193	1193	1193
45	1194	1194	1194
46	1195	1195	1195
47	1196	1196	1196
48	1197	1197	1197
49	1198	1198	1198
50	1199	1199	1199
51	1200	1200	1200
52	1201	1201	1201
53	1202	1202	1202
54	1203	1203	1203
55	1204	1204	1204
56	1205	1205	1205
57	1206	1206	1206
58	1207	1207	1207
59	1208	1208	1208
60	1209	1209	1209
61	1210	1210	1210
62	1211	1211	1211
63	1212	1212	1212
64	1213	1213	1213
65	1214	1214	1214
66	1215	1215	1215
67	1216	1216	1216
68	1217	1217	1217
69	1218	1218	1218
70	1219	1219	1219
71	1220	1220	1220
72	1221	1221	1221
73	1222	1222	1222
74	1223	1223	1223
75	1224	1224	1224
76	1225	1225	1225
77	1226	1226	1226
78	1227	1227	1227
79	1228	1228	1228
80	1229	1229	1229
81	1230	1230	1230
82	1231	1231	1231
83	1232	1232	1232
84	1233	1233	1233
85	1234	1234	1234
86	1235	1235	1235
87	1236	1236	1236
88	1237	1237	1237
89	1238	1238	1238
90	1239	1239	1239
91	1240	1240	1240
92	1241	1241	1241
93	1242	1242	1242
94	1243	1243	1243
95	1244	1244	1244
96	1245	1245	1245
97	1246	1246	1246
98	1247	1247	1247
99	1248	1248	1248
100	1249	1249	1249
101	1250	1250	1250
102	1251	1251	1251
103	1252	1252	1252
104	1253	1253	1253
105	1254	1254	1254
106	1255	1255	1255
107	1256	1256	1256
108	1257	1257	1257
109	1258	1258	1258
110	1259	1259	1259
111	1260	1260	1260
112	1261	1261	1261
113	1262	1262	1262
114	1263	1263	1263
115	1264	1264	1264
116	1265	1265	1265
117	1266	1266	1266
118	1267	1267	1267
119	1268	1268	1268
120	1269	1269	1269
121	1270	1270	1270
122	1271	1271	1271
123	1272	1272	1272
124	1273	1273	1273
125	1274	1274	1274
126	1275	1275	1275
127	1276	1276	1276
128	1277	1277	1277
129	1278	1278	1278
130	1279	1279	1279
131	1280	1280	1280
132	1281	1281	1281
133	1282	1282	1282
134	1283	1283	1283
135	1284	1284	1284
136	1285	1285	1285
137	1286	1286	1286
138	1287	1287	1287
139	1288	1288	1288
140	1289	1289	1289
141	1290	1290	1290
142	1291	1291	1291
143	1292	1292	1292
144	1293	1293	1293
145	1294	1294	1294
146	1295	1295	1295
147	1296	1296	1296
148	1297	1297	1297
149	1298	1298	1298
150	1299	1299	1299
151	1300	1300	1300
152	1301	1301	1301
153	1302	1302	1302
154	1303	1303	1303
155	1304	1304	1304
156	1305	1305	1305
157	1306	1306	1306
158	1307	1307	1307
159	1308	1308	1308
160	1309	1309	1309
161	1310	1310	1310
162	1311	1311	1311
163	1312	1312	1312
164	1313	1313	1313
165	1314	1314	1314
166	1315	1315	1315
167	1316	1316	1316
168	1317	1317	1317
169	1318	1318	1318
170	1319	1319	1319
171	1320	1320	1320
172	1321	1321	1321
173	1322	1322	1322
174	1323	1323	1323
175	1324	1324	1324
176	1325	1325	1325
177	1326	1326	1326
178	1327	1327	1327
179	1328	1328	1328
180	1329	1329	1329
181	1330	1330	1330
182	1331	1331	1331
183	1332	1332	1332
184	1333	1333	1333
185	1334	1334	1334
186	1335	1335	1335
187	1336	1336	1336
188	1337	1337	1337
189	1338	1338	1338
190	1339	1339	1339
191	1340	1340	1340
192	1341	1341	1341
193	1342	1342	1342
194	1343	1343	1343
195	1344	1344	1344
196	1345	1345	1345
197	1346	1346	1346
198	1347	1347	1347
199	1348	1348	1348
200	1349	1349	1349
201	1350	1350	1350
202	1351	1351	1351
203	1352	1352	1352
204	1353	1353	1353
205	1354	1354	1354
206	1355	1355	1355
207	1356	1356	1356
208	1357	1357	1357
209	1358	1358	1358
210	1359	1359	1359
211	1360	1360	1360
212	1361	1361	1361
213	1362	1362	1362
214	1363	1363	1363
215	1364	1364	1364
216	1365	1365	1365
217	1366	1366	1366
218	1367	1367	1367
219	1368	1368	1368
220	1369	1369	1369
221	1370	1370	1370
222	1371	1371	1371
223	1372	1372	1372
224	1373	1373	1373
225	1374	1374	1374
226	1375	1375	1375
227	1376	1376	1376
228	1377	1377	1377
229	1378	1378	1378
230	1379	1379	1379
231	1380	1380	1380
232	1381	1381	1381
233	1382	1382	1382
234	1383	1383	1383
235	1384	1384	1384
236	1385	1385	1385
237	1386	1386	1386
238	1387	1387	1387
239	1388	1388	1388
240	1389	1389	1389
241	1390	1390	1390
242	1391	1391	1391
243	1392	1392	1392
244	1393	1393	1393
245	1394	1394	1394
246	1395	1395	1395
247	1396	1396	1396
248	1397	1397	1397
249	1398	1398	1398
250	1399	1399	1399
251	1400	1400	1400
252	1401	1401	1401
253	1402	1402	1402
254	1403	1403	1403
255	1404	1404	1404
256	1405	1405	1405
257	1406	1406	1406
258	1407	1407	1407
259	1408	1408	1408
260	1409	1409	1409
261	1410	1410	1410
262	1411	1411	1411
263	1412	1412	1412
264	1413	1413	1413
265	1414	1414	1414
266	1415	1415	1415
267	1416	1416	1416
268	1417	1417	1417
269	1418	1418	1418
270	1419	1419	1419
271	1420	1420	1420
272	1421	1421	1421
273	1422	1422	1422
274	1423	1423	1423
275	1424	1424	1424
276	1425	1425	1425
277	1426	1426	1426
278	1427	1427	1427
279	1428	1428	1428
280	1429	1429	1429
281	1430	1430	1430
282	1431	1431	1431
283	1432	1432	1432
284	1433	1433	1433
285	1434	1434	1434
286	1435	1435	1435
287	1436	1436	1436
288	1437	1437	1437
289	1438	1438	1438
290	1439	1439	1439
291	1440	1440	1440
292	1441	1441	1441
293	1442	1442	1442
294	1443	1443	1443
295	1444	1444	1444
296	1445	1445	1445
297	1446	1446	1446
298	1447	1447	1447
299	1448	1448	1448
300	1449	1449	1449
301	1450	1450	1450
302	1451	1451	1451
303	1452	1452	1452
304	1453	1453	1453
305	1454	1454	1454
306	1455	1455	1455
307	1456	1456	1456
308	1457	1457	1457
309	1458	1458	1458
310	1459	1459	1459
311	1460	1460	1460
312	1461	1461	1461
313	1462	1462	1462
314	1463	1463	1463
315	1464	1464	146

XX Claim 21; Fig 5C; 66pp; English.
PS The sequence represents the coding sequence of cardiac enhancer
XX hcsx/Nkx2.5 homology domain B. The nucleic acid is useful for
CC specifically expressing a gene in a cardiac cell, as an earlier marker
CC of cardiomyocyte induction, e.g. for optimising cardiomyocyte induction.
CC Genes expressed in the cardiac cell-specific manner are useful for the
CC targeted expression of genes encoding therapeutic proteins for the
CC treatment of damaged heart tissue. Cardiac specific enhancer elements may
CC be used for gene therapy.
XX
XX Sequence 478 BP; 138 A; 129 C; 89 G; 122 T; 0 other;

Query Match 100.0%; Score 478; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.9e-122;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agagaatcattaccgattccacaaagacatagagagttaacagtcactgattgtt 60
Db 1 agagaatcattaccgattccacaaagacatagagagttaacagtcactgattgtt 60
Qy 61 caaataggagagtttttttcccttccctttttgtaacacctgacccacagactgacag 120
Db 61 caaataggagagtttttttcccttccctttttgtaacacctgacccacagactgacag 120
Qy 121 ttctagaagcccttaccgaaataggaataaataacattccttgccacctgatttgcag 180
Db 121 ttctagaagcccttaccgaaataggaataaataacattccttgccacctgatttgcag 180
Qy 181 ggcaatgtaatttttttctccagagctctcaaaaaaaaccccttac 240
Db 181 ggcaatgtaatttttttctccagagctctcaaaaaaaaccccttac 240
Qy 241 taaaaacaggatcccgatgtagcctgatgtcccccattaaacggttaatttcaggc 300
Db 241 taaaaacaggatcccgatgtagcctgatgtcccccattaaacggttaatttcaggc 300
Qy 301 gtccgctcacactaattttcaaaactgtcatcgagcgccctggccagcagattcactt 360
Db 301 gtccgctcacactaattttcaaaactgtcatcgagcgccctggccagcagattcactt 360
Qy 361 aacagcgctcccgagacctgttccgagctctttccagcgagacatttaattgaatcgg 420
Db 361 aacagcgctcccgagacctgttccgagctctttccagcgagacatttaattgaatcgg 420
Qy 421 atgtgctcgtttccagacgtcaccgctcgcgataggcatctctccacagacac 478
Db 421 atgtgctcgtttccagacgtcaccgctcgcgataggcatctctccacagacac 478

RESULT 2
AAS09962
ID AAS09962 standard; DNA; 6751 BP.
XX
XX AAS09962;
AC
XX
DT 24-OCT-2001 (first entry)
XX
DE Genomic DNA #2 encoding human Csx/Nkx2.5.
XX
XX Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
KW therapeutic; heart tissue; gene therapy; human; ds.
XX
OS Homo sapiens.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX

PR 14-JAN-2000; 2000US-0176419.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA
XX Lee IW, Izumo S;
XX WPT; 2001-451809/48.
DR
XX New cardiac specific cell enhancer elements, useful for specifically
XX expressing gene in cardiac cell, as earlier marker of cardiomyocyte
XX induction, e.g. for optimizing cardiomyocyte induction -
PS Claim 29; Fig 4B; 66pp; English.
XX
XX The sequence represents the genomic sequence #2 of human Csx/Nkx2.5. The
XX nucleic acid is useful for specifically expressing a gene in a cardiac
XX cell, as an earlier marker of cardiomyocyte induction, e.g. for
XX optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
XX specific manner are useful for the targeted expression of genes encoding
XX therapeutic proteins for the treatment of damaged heart tissue. Cardiac
XX specific enhancer elements may be used for gene therapy.
SQ Sequence 6751 BP; 1459 A; 2008 C; 1753 G; 1490 T; 41 other;

Query Match 100.0%; Score 478; DB 22; Length 6751;
Best Local Similarity 100.0%; Pred. No. 5.1e-122;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agagaatcattaccgattccacaaagacatagagagttaacagtcactgattgtt 60
Db 201 agagaatcattaccgattccacaaagacatagagagttaacagtcactgattgtt 260
Qy 61 caaataggagagtttttttcccttccctttttgtaacacctgacccacagactgacag 120
Db 261 caaataggagagtttttttcccttccctttttgtaacacctgacccacagactgacag 320
Qy 121 ttctaggagcccttaccgaaataggaataaataacattccttgccacctgatttgcag 180
Db 321 ttctaggagcccttaccgaaataggaataaataacattccttgccacctgatttgcag 380
Qy 181 ggcaatgtaatttttttctccagagctctcaaaaaaaaccccttac 240
Db 381 ggcaatgtaatttttttctccagagctctcaaaaaaaaccccttac 440
Qy 241 taaaaacaggatcccgatgtagcctgatgtcccccattaaacggttaatttcaggc 300
Db 441 taaaaacaggatcccgatgtagcctgatgtcccccattaaacggttaatttcaggc 500
Qy 301 gtccgctcacactaattttcaaaactgtcatcgagcgccctggccagcagattcactt 360
Db 501 gtccgctcacactaattttcaaaactgtcatcgagcgccctggccagcagattcactt 560
Qy 361 aacagcgctcccgagacctgttccgagctctttccagcgagacatttaattgaatcgg 420
Db 561 aacagcgctcccgagacctgttccgagctctttccagcgagacatttaattgaatcgg 620
Qy 421 atgtgctcgtttccagacgtcaccgctcgcgataggcatctctccacagacac 478
Db 621 atgtgctcgtttccagacgtcaccgctcgcgataggcatctctccacagacac 678

RESULT 3
AAC59534
ID AAC59534 standard; cDNA; 1086 BP.
XX
XX AAC59534;
AC
XX
DT 15-FEB-2001 (first entry)
XX
DE Human secreted protein cDNA sequence #28.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;


```
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT respiratory disorders related to the lung including lung cancers and
PT also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID NO 362; 507pp; English.
XX
XX Sequences AAS29931-AA530164 represent genomic DNA molecules, which encode
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides
CC and their associated polynucleotides are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by detecting the presence or
CC absence of a mutation in a lung antigen polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 8.2%; Score 39; DB 22; Length 3276;
XX Best Local Similarity 68.4%; Pred. No. 0.85;
XX Matches 54; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
XX
XX 177 caaggcgcaatgctaatattttttcttctccagagctctcaaaaaaataaaacc 236
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2059 catggacaatgaattttttcttcttaccctgttaaaaaaataaaacc 2118
XX
XX 237 ttactaaaaacagggtacc 255
XX ||||| ||||| |||||
XX 2119 taataaaaaaaaagtc 2137
XX
XX RESULT 6
XX AA193585/c
XX ID AA193585 standard; cDNA; 397 BP.
XX
XX AA193585;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 13645.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
```

```
KW nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200164835-A2.
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO13654.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 1; SEQ ID NO 13645; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AA193841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 397 BP; 115 A; 95 C; 72 G; 115 T; 0 other;
XX
XX Query Match 8.0%; Score 38.4; DB 22; Length 397;
XX Best Local Similarity 56.2%; Pred. No. 0.57;
XX Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
XX
XX 121 ttctaggagcccttaccgcgaataataggaaataatctcttgccaccttgattgcaag 180
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 208 TTTTGTGGGCGCCGCAAAAAAATAAAAAAACCCTCTTATTTTGGGGG 149
XX
XX 181 ggcaatgctaattttttcttctccagagctctcaaaaaaataaaacccttac 240
XX ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 148 GGGGCCCCCTTTTCTTTTTCCTCCCTCCCAACAAAAAATAAAAAA 89
XX
XX 241 taataaaca 248
XX |||||
XX 88 AAAAAAAA 81
XX
XX RESULT 7
XX AA191430
XX ID AA191430 standard; cDNA; 403 BP.
XX
XX AA191430;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 11490.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
```



```

PN W0200116046-A2.
XX
PU 15-MAR-2001.
XX
PF 06-SEP-2000; 2000W0-US24827.
XX
PR 10-SEP-1999; 990S-0394374.
PR 01-MAY-2000; 2000US-0561778.
PR 15-AUG-2000; 2000US-0640173.
PR 07-SEP-2000; 2000US-0656668.
XX
XX {CORT-} CORIXA CORP.
XX
XX Xu J, Stolk JA;
XX
XX WPI; 2001-211395/21.
XX
XX Isolated polypeptides associated with ovarian carcinomas, and the
XX nucleic acids that encode them, useful for the prevention diagnosis and
XX treatment of ovarian cancers -
XX
XX Claim 5; Page 126; 189pp; English.
XX
XX The present invention provides a number of coding sequences and proteins,
XX the over-expression of which is associated with ovarian carcinoma/cancer.
XX These can be used in the diagnosis, treatment and prevention of ovarian
XX cancer, optionally by gene therapy or in the form of a vaccine. The
XX present sequence is an example of one of these sequences.
XX
XX Sequence 396 BP; 60 A; 17 C; 34 G; 162 T; 123 other;
SU

Query Match      8.0%; Score 38.2; PR 22; Length 396;
Best Local Similarity 37.2%; Pred. No. 0.65;
Matches 64; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 77 ttttctctcccttttttaacacctgaacccacagagactgaactctagaagacccct 136
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 259 TTTTTCNNNNNNNANCCGCCGNCNCCNNNGGGNTTTNNNNNANNNNAANTN 200

QY 137 tacccgaataataataatctctgccacctgaatttcaagagcaactaatatt 196
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 199 NTAAAGACCCGCGNAAANNNNNNNNNNNNTTTTNNNNNNNTTTTTTTTTT 140

QY 197 tttcttctccagagctctcaaaaaaaataaaaaaaacacttactaaaaaca 248
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 139 TTNNTNNNNNANNNCNNAANAAAAAANAAAAAANAAAAAANAAAAA 88

RESULT 10
AAK64788/c
ID AAK64788 standard; DNA: 10740 BP.
XX
XX AAK64788;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19600.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX W0200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001W0-US01354.
XX
XX 31-JAN-2000; 2000US-0175055.
PR
PR 04-FEB-2000; 2000US-0180628.
PR
PR 24-FEB-2000; 2000US-0184564.
PR
```

```

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214889.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216890.
PR 31-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220566.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225347.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229348.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229518.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231318.
PR 08-SEP-2000; 2000US-0231314.
PR 08-SEP-2000; 2000US-0231314.
PR 08-SEP-2000; 2000US-0232083.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234947.
PR 25-SEP-2000; 2000US-0234948.
PR 26-SEP-2000; 2000US-0235384.
PR 27-SEP-2000; 2000US-0235814.
PR 27-SEP-2000; 2000US-0235816.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR
```


XX 15-MAR-2001; 2001WO-EP02945.
PF 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602751/68.
XX Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX Claim 1; SEQ ID NO 185; 28pp; English.
XX Sequences AAS45296-AAS4520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;

Query Match 7.9%; Score 37.8; DB 22; Length 11047;
Best Local Similarity 51.1%; Pred. No. 2.9;
Matches 115; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
Qy 24 aaagacatagagagtgaacagtcactgtctgttcacaaataggaggaggttttttcc 83
Db 7544 AAAAAAAAAAATAATACATCAATTAATTAATAAATAATTTATACCTTTTAACC 7485
Qy 84 ttccctttttgttaa---cacctgacccacagagactgtctctaggagcccccctacc 140
Db 7484 TCCATTATTACGAATTTCAATTCACCAATAAATCAATAATATTATAGGCACCTACTTCT 7425
Qy 141 cgaaaataggaaataaactcttgccaccttgatttgcgaagggaatgctaatttttct 200
Db 7424 TCTAAATCTTCATTAATAACTACTAACCATATACTAAATCGCTATTTCCTCAAAATTTAAAT 7365
Qy 201 ttctccagagctctcaaaaaaaaaaaaaaaaaaaacccctactcaaaa 245
Db 7364 ATTTCTCACTATATAAAAAAAAAAAAAAAAAATAAATAATACTTAA 7320

RESULT 14
ABL33985/C
ID ABL33985 standard; DNA; 11047 BP.
XX ABL33985;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1958.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX Homo sapiens.
OS WO200200928-A2.
PN 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX Claim 1; SEQ ID NO 1958; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulceration bowel
CC diseases. The present sequence is a gene of the invention.
XX Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;

Query Match 7.9%; Score 37.8; DB 24; Length 11047;
Best Local Similarity 51.1%; Pred. No. 2.9;
Matches 115; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
Qy 24 aaagacatagagagtgaacagtcactgtctgttcacaaataggaggaggttttttcc 83
Db 7544 AAAAAAAAAAATAATACATCAATTAATTAATAAATAATTTATACCTTTTAACC 7485
Qy 84 ttccctttttgttaa---cacctgacccacagagactgtctctaggagcccccctacc 140
Db 7484 TCCATTATTACGAATTTCAATTCACCAATAAATCAATAATATTATAGGCACCTACTTCT 7425
Qy 141 cgaaaataggaaataaactcttgccaccttgatttgcgaagggaatgctaatttttct 200
Db 7424 TCTAAATCTTCATTAATAACTACTAACCATATACTAAATCGCTATTTCCTCAAAATTTAAAT 7365
Qy 201 ttctccagagctctcaaaaaaaaaaaaaaaaaaaacccctactcaaaa 245
Db 7364 ATTTCTCACTATATAAAAAAAAAAAAAAAAAATAAATAATACTTAA 7320

RESULT 15
AAH76837
ID AAH76837 standard; cDNA; 5027 BP.
XX AAH76837;
XX 14-DEC-2001 (first entry)
XX Human vesicular transport-related protein 11-encoding cDNA.

Human; vesicular transport-related protein 11; recombinant production; malignant tumour; cancer; blood disease; HIV infection; human immunodeficiency virus; immune disorder; inflammatory condition; cytostatic; anti-HIV; antiinflammatory; immunomodulator; ss.	Human sapiens.
Key	Location/Qualifiers
CDS	4448..4756
FT	/xref= a
FI	/product= "human vesicular transport-related protein 11"
XX	W:200174816-A1.
PN	04-OCT-2001.
PU	
PF	26-MAR-2001; 2001WO-CN00493.
PP	27-MAR-2000; 2000CN-0115175.
PR	(SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
PA	
XX	Mao Y, Xie Y;
PI	WFI: 2001-602451/48.
XX	P-USUB: AAG5760.
XX	Human vesicular transport-related protein 11 and encoded polynucleotide, applicable in diagnosis and treatment of malignant tumour, hemopathy, HIV infection, immunological diseases and various inflammations
XX	Claim 6: Page 29-32; 36pp: Chinese.
XX	This sequence represents cDNA encoding human vesicular transport-related protein 11. The protein has a molecular weight of 11 kD. The invention relates to human vesicular transport-related protein 11 (AAG566760), nucleic acids encoding it (AAG5837), and a method for the recombinant production of vesicular transport-related protein 11. The present invention additionally discloses an antagonist of vesicular transport- related protein 11 for therapeutic use, and an antibody which specifically binds to vesicular transport-related protein 11. Vesicular transport-related protein 11, and nucleotides which encode it may be used for treating a variety of diseases, such as malignant tumours, blood diseases, HIV (human immunodeficiency virus) infection, immune disorders and inflammatory conditions. The protein may also be used to screen for modulators of its activity or for peptide fingerprinting identification. The polynucleotide can be used as a primer for nucleic acid amplification reactions or as a probe for hybridisation reactions, or in producing gene chips or microarrays.
XX	Sequence 5027 BP: 1811 A; 760 G; 735 G; 1721 T; 0 other;

Query Match	7.88;	Score 37.4;	DB 22;	Length 5027;
Best local Similarity	46.7%;	Pred. No. 2.8;		
Matches 119;	Conservative 0;	Mismatches 136;	Indels 0;	Gaps 0;

QY	30	catagagagatgaacagtcactgactctgtgttcacaaatagagagagctttttttcttccct	89
Db	1618	cgttgcatttaagtgtcaaaatttgattttctataagagatgtcactcttagctatcact	1677
QY	90	tttttttaacacctgaacccacagagacttgcattctctagaaagccctttaccgagaaata	149
Db	1678	ttttgtgattgctaatgagatgataatgataatctcaagagataatttgcacttagattt	1737
QY	150	gaataaatactgtccacacctgatttgcagagagcaactgctaatttttttcttctccaga	209
Db	1738	ggnataatgctctcttttatacattcaaaagagataaattccacactctcttcagagct	1797
QY	210	gctctcaaaaaaataaaaaaaaccttactcaaaaagagatcccgaatgagctga	249
Db	1798	ataaacacacaaacagaaaaaacacctttgcagagaaactaaagatgcaatttctctttt	1857


```

: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SLE
: LOCATION: (1033)
: OTHER INFORMATION: n equals a,t,c, or c
US-92-227-357-40

```

```

Query Match: 4.94%  TP 4.1  Length 1059;
Best Local Similarity 57.1%;  Pred. No. 1.5;
Matches 60;  Conservative 0;  Mismatches 47;  Gaps 0;  Gaps
QY 144 aaatagaataaataactcttccagctttatctatcaatgagctactgaattttctctctc 205
Db 951 AIAIINGAGTCGAAACAAACATATATATTAIAGGAGATTTCTTCTTTTCATTT 1111
QY 204 tccagagcctctcaaaataaaataaaataaaataaaataaaataaaataaa 248
Db 1011 TAAAAAATCTCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1055

RESULT 6
PCT-DS94 -14(073-1
: Sequence 1, Application: PCT/US-414073
: GENERAL INFORMATION:
: APPLICANT: THOMAS, WAYNE B., CHA, KONG-YAN, POKES, BRUCE L., and
: APPLICANT: KIM, MINGJAE
: TITLE OF INVENTION: NOVEL ACTING ENZYME A HOUSEHOLD MICE
: TITLE OF INVENTION: ALDERGEN, DEEP FILL ANALYSIS THEREFOR
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lohmeier & Lockhold
: STREET: 60 State Street, Suite 510

```

```
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14073
; FILING DATE: 08-DEC-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-072 (IMI-041)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 63..848
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 150..848
; PCT-US94-14073-1

Query Match 6.9%; Score 33; DB 5; Length 1059;
Best Local Similarity 57.1%; Pred. No. 1.5;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 144 aaatagagaaataatccttgcacccttgatttgcagggaatgctaatctttttcttc 203
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 951 ATATTGAGCTGAAACAAACAAATCATGATTGATAGGATTTCCTTGTGTTTCAATT 1010
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 204 tccagagctctcaaaaaaaaaaaaaaaaaaaccttactaaaaa 248
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1011 TAAAAAATCCAAAATAAATAATAATAATCATTTCAACAAAAA 1055
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-08-487-810-1
; Sequence 1, Application US/08487810
; Patent No. 5618695
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA ENCODING HEM-1, A GENE EXPRESSED BY
; TITLE OF INVENTION: SCLEROSING HEMANGIOMA CELLS
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-487-810-1

Query Match 6.9%; Score 32.8; DB 1; Length 1146;
Best Local Similarity 59.8%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 143 aaatagagaaataatccttgcacccttgatttgcagggaatgctaatctttttcttc 202
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1054 AAAATCGCAGTTAAAGCATGAACACATTGAATTGAAAAAGAACTCTGTAGTTTGAGATT 1113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 203 ctcagagctctcaaaaaaaaaaaaaaaaaaaaaa 234
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1114 TGCCATACAGCAAGAAAAAATAAAAAA 1145
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-08-553-367A-5
; Sequence 5, Application US/08553367A
; Patent No. 5939539
; GENERAL INFORMATION:
; APPLICANT: Theodor LANGE et al.
; TITLE OF INVENTION: GA 20-OXIDASE GENE SEQUENCES (As Amended)
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,367A
; FILING DATE: No. 5939539ember 27, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 49/FD4.5WZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; IMMEDIATE SOURCE:
; CLONE: PAT2353
; US-08-553-367A-5

Query Match 6.9%; Score 32.8; DB 2; Length 1490;
Best Local Similarity 59.8%; Pred. No. 2;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```



```
QY 216 aaaaataaaaaaaacccctactaaaa 245
|||||
Db 1471 AAAAAAAAAAAAAACCCTATTTTCAAA 1500

RESULT 13
US-08-860-334-17
: Sequence 17, Application US/08860339
: Patent No. 6117665
: GENERAL INFORMATION:
: APPLICANT: Kossmann, Jens
: APPLICANT: Emmertmann, Michael
: APPLICANT: Virgin, Ivar
: APPLICANT: Renz, Andreas
: TITLE OF INVENTION: DNA MOLECULES CODING FOR DEBRANCHING ENZYMES DERIVED
: TITLE OF INVENTION: FROM PLANTS
: FILE REFERENCE: AGREVO-6
: CURRENT APPLICATION NUMBER: US/08/860,339
: EARLIER FILING DATE: 1997-11-25
: EARLIER APPLICATION NUMBER: DE P4447487.7
: EARLIER FILING DATE: 1994-12-22
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 17
: LENGTH: 3437
: TYPE: DNA
: ORGANISM: Spinacia oleracea
: FEATURE:
: NAME/KEY: CUS
: LOCATION: (20)..(3095)
US-08-860-334-17

Query Match 6.8%; Score 32.4; DB 3; Length 3437;
Best Local Similarity 54.1%; Pred. No. 3.6;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 127 qaaccrccctaccgaataaataaataaacttttgcaccttgatttcaaggagcaat 186
|||||
Db 3415 ggaatgctttgttcgcgcataataacttctgttgaaccatacatcctatgaat 3374

QY 187 gtaataattttttctctccagagctctcaaaaaaataaaaaaaacccctactaaaaa 246
|||||
Db 3375 qaataatttttttttttaaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 3434

QY 247 ca 248
Db 3435 aa 3436

RESULT 14
US-09-248-335-57
: Sequence 57, Application US/09248335
: Patent No. 6096504
: GENERAL INFORMATION:
: APPLICANT: MCGUNIGLE, BRIAN
: APPLICANT: O'KEEF, DANIEL
: TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
: FILE REFERENCE: CL-1128-A
: CURRENT APPLICATION NUMBER: US/09/248,335
: CURRENT FILING DATE: 1999-02-10
: EARLIER APPLICATION NUMBER: 08/924,759
: EARLIER FILING DATE: 1997-September-05
: NUMBER OF SEQ ID NOS: 74
: SOFTWARE: Microsoft Word Version 7.0A
: SEQ ID NO 57
: LENGTH: 960
: TYPE: DNA
: ORGANISM: maize
US-09-248-335-57
```

```
Query Match 6.7%; Score 31.6; DB 3; Length 960;
Best Local Similarity 64.0%; Pred. No. 3.2;
Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 174 tturaaaacaaatcaatatttttttttttccaaatctcaaaataaaaaaataaaa 243
|||||
Db 856 ttataatccaatctaatctatctctatctatctatctatctatctatctatctat 244

QY 244 accttactaaaaaa 248
|||||
Db 916 aaaaaaataaaaaaa 930

RESULT 15
US-08-969-106-9
: Sequence 9, Application US/08969106
: Patent No. 5986055
: GENERAL INFORMATION:
: APPLICANT: Yang, M.
: APPLICANT: Nandabalan, K.
: APPLICANT: Schulz, V.
: TITLE OF INVENTION: CLK2 INTERACTIONS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Petrule & Edwards LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10044
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/969,106
: FILING DATE: 13-NOV-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MISROCK, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/POCKET NUMBER: 7934-057
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-794-9090
: TELEFAX: 212 859-9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2399 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
US-08-969-106-9

Query Match 6.6%; Score 31.6; DB 2; Length 2399;
Best Local Similarity 53.2%; Pred. No. 5.4;
Matches 67; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 109 caaaactcaaatctctataaaaaaacccctactcaaaataataaaataaactctaccacc 149
|||||
Db 2260 CAGGACACCGCIGGAGTAGCTTCGCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCT 2314

QY 169 ttcaatttcaaaatcaatgaatatttttttttttttttttttttttttttttttttt 228
|||||
Db 2320 TTGAGATTGAGTGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2479

QY 229 aaaaaa 244
|||||
Db 2380 AAATAA 2385
```

Search completed: July 5, 2002, 08:06:25
Job time: 6517 sec

Genome version 4.5
Copyright (c) 1993 - 2000 Computer Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 06:11:43 : Search time 1604.3 Seconds
(without alignments)
4021.408 Million cell updates/sec

Title: us-09-761-466-6

Perfect score: 478
Sequence: 1 aagaaatcatcaccgatt.....ggaatcctctcaaacacac 478

Scoring table: GENETIC-NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 5748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 7

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100

Maximum Match 100

Listed first 45 summaries

Database :

EST :

1: em_est1a1*
2: em_est1b1*
3: em_est1c1*
4: em_est1d1*
5: em_est1e1*
6: em_est1f1*
7: em_est1g1*
8: em_est1h1*
9: em_est1i1*
10: em_est1j1*
11: em_est1k1*
12: em_est1l1*
13: em_est1m1*
14: em_est1n1*
15: em_est1o1*
16: em_est1p1*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Res ID	Score	Query Match	Length	DB ID	Description
C 1	45.6	9.5	1101	12	CNS014PM
C 2	43.6	9.1	838	10	EG781725 SEAM3001
C 3	43.2	9.0	1244	12	CNS014V
C 4	41.8	8.7	453	12	CNS014V
C 5	41.8	8.7	1101	12	CNS0019R
C 6	41.6	8.7	289	9	A1947755
C 7	41.6	8.7	981	9	A1947755
C 8	41	8.6	791	12	CNS014PQ
C 9	40.8	8.5	829	9	AL514053
C 10	40.8	8.5	1201	12	CNS014V
C 11	40.6	8.5	367	10	B1540842
C 12	40.2	8.4	859	12	CNS004Y
C 13	40.2	8.4	980	12	CNS000Y
C 14	40.2	8.4	986	12	CNS000Y
C 15	40.4	8.4	1101	12	CNS00AY1
C 16	40	8.4	267	10	C41850
C 17	40	8.4	343	9	AL355572

C 18	40	8.4	1101	12	CNS014V
C 19	40	8.4	1543	10	B152111
C 20	40.8	8.3	528	12	AG000000
C 21	40.8	8.3	657	12	CNS014V
C 22	40.8	8.3	906	12	CNS014V
C 23	40.6	8.3	450	12	AG000000
C 24	40.6	8.3	529	10	BF200012
C 25	40.6	8.3	540	12	AG000000
C 26	40.4	8.2	1101	12	CNS000P
C 27	40.2	8.2	1019	10	BF741517
C 28	39	8.2	520	9	A1701478
C 29	39	8.2	1101	12	CNS014V
C 30	39	8.2	1201	12	CNS014V
C 31	38.8	8.1	352	9	AL500444
C 32	38.8	8.1	466	9	AL500444
C 33	38.8	8.1	902	12	CNS000P
C 34	38.6	8.1	166	9	A171214
C 35	38.6	8.1	253	9	A1829472
C 36	38.4	8.0	653	10	B6247000
C 37	38.4	8.0	880	9	AL521604
C 38	38.4	8.0	1101	12	CNS014V
C 39	38.4	8.0	1101	12	CNS014V
C 40	38.2	8.0	268	10	BF701447
C 41	38.2	8.0	272	10	BF701447
C 42	38.2	8.0	295	10	BF701447
C 43	38.2	8.0	422	9	AW470444
C 44	38.2	8.0	493	9	AW470444
C 45	38.2	8.0	600	9	A1472000

ALIGNMENTS

RESULTS :
CNS014PM/1
E-VALUE
DEFINITION
CNS014PM/1
Prosophila melanogaster genomic survey, sequence 17, end of PA
RAC1F614 of Drosophila melanogaster genomic survey sequence
1133 genomic survey sequence

ALIGNMENT
AL107044
AL107044.1 31:56,25157

VERSION :
GENES

KEYWORDS
fruit fly,

ORGANISM
Drosophila melanogaster

Prokaryote: Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;

Protista; Neoptera; Eukaryota; Eukaryota; Insecta; Insecta;

Muscomorpha; Ephydroidea; Insecta; Insecta; Insecta;

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (28-JUL-1999)

BP 191 91006 EVRY cedex - FRANCE (E-mail : sequenc@genoscope.fr)

Web : www.genoscope.cns.fr

Determination of this PAC-seq sequence was carried out as part of a

collaboration with the European Institute of Science and Technology (EIST)

Library (EIST PAC) was made by Anne B. and Anne B. (EIST)

Project grant. The DNA was prepared from cells by a PCR

and GeneView Payan. It has been constructed in the vector

phelobac11.

FEATURES

Location/Qualities

1..1101

Organization: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"

Feature: "Drosophila melanogaster"


```
Qy 118 cagttctaggaagccctaccgaaataggaaataatccttgccacctgtattgc 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CAGTTTACGGGTGTCGCCCATCCACCACATGATAATATAGTTTGCACAATTATTGCTC 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 178 aagggaatgctaattttttttcttcagagagctctcaaaaaaaaaaaaaaaaaac 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 AGACGCTTCCCAATCGAAGACTTTTCTGCCATGCTTAAAAAaaaaaaaaaaaaa 285
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS AI947756 381 bp mRNA linear EST 19-AUG-1999
DEFINITION 603027H09.xl 603 - stressed root cDNA library from Wang/Bohnert lab
Zea mays cDNA, mRNA sequence.
ACCESSION AI947756
VERSION AI947756.1 GI:5740066
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 381)
REFERENCE Walbot,V.
AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 603027 row: H column: 09.
FEATURES
source
1..381
Location/Qualifiers
1..381
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/db_lib="603 - stressed root cDNA library from
Wang/Bohnert lab"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E. coli XL Gold"
/note="Organ: root; Vector: pBluescriptII SK(+); XR;
Seedling stressed root cDNA library from Wang/Bohnert lab"
BASE COUNT 122 a 72 c 84 g 103 t
ORIGIN

Query Match 8.7%; Score 41.6; DB 9; Length 381;
Best Local Similarity 59.2%; Pred. No. 3.7e+02;
Matches 71; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 118 cagttctaggaagccctaccgaaataggaaataatccttgccacctgtattgc 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 CAGTTTACGGGTGTCGCCCATCCACCACATGATAATATAGTTTGCACAATTATTGCTC 317
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 178 aagggaatgctaattttttcttcagagagctctcaaaaaaaaaaaaaaaaaac 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 AGACGCTTCCCAATCGAAGACTTTTCTGCCATGCTTAAAAAaaaaaaaaaaaaa 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
LOCUS CNS04DPQ/c 791 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
103G11 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL286055
VERSION AL286055.1 GI:8024497
KEYWORDS GSS; genome survey sequence.
```

```
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 791)
REFERENCE Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
AUTHORS Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 791)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 791)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..791
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="103G11"
/clone_lib="G"
/note="Genoscope sequence ID : COBG103AD06SP1-end ;
PUC-ori"
BASE COUNT 208 a 146 c 144 g 273 t 20 others
ORIGIN

Query Match 8.6%; Score 41; DB 12; Length 791;
Best Local Similarity 48.6%; Pred. No. 3.2e+02;
Matches 88; Conservative 8; Mismatches 85; Indels 0; Gaps 0;

Qy 136 ttaccgaaataggaaataatccttgccacctgtattgcaaggcgaatgctaatttt 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 TTTACCTHWKTKTTGCGADAACCTCTCATCTCTAATTTGKAATCAGCCWTTTTT 560
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 196 tttctttccagagctctcaaaaaaaaaaaaaaaaaacaccttactaaaaacagggtacc 255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 559 TTTCTTTTCAACCCCTTTAAAAAaaaaaaaaaaaaaaaaaatacttACAGCTTGAATTC 500
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 256 cggatgtagctcgatgtcccccattaaacggtaatatttcaggcgctccacacataa 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 499 AAACAGTTGGAGCTGAGGTGTTTCTTACCTTTGACATGAGACAAATCCCAAGCAA 440
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 316 t 316
Db 439 T 439

RESULT 9
LOCUS AL514053/c 829 bp mRNA linear EST 13-FEB-2001
DEFINITION AL514053 LTI_NFL006.PL2 Homo sapiens cDNA clone CL0BA012G07 3
prime, mRNA sequence.
ACCESSION AL514053
VERSION AL514053.1 GI:12777547
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```


REFERENCE 1 (bases 1 to 829)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact : Genoscope
Genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES source location/Qualifiers
1..829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CL0BA012G07"
/tissue_type="PLT.NFL006.PL2"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng liang life technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 616 8371 Email : liliang@lifetech.com URL : http://tullnberg.invitrogen.com"

BASE COUNT 213 a 153 g 103 q 247 t 113 others

ORIGIN

Query Match 8.5%; Score 40.8; DB 9; Length 829;
Best local Similarity 42.4%; Prod. No. 3,3e+02;
Matches 84; Conservative 22; Mismatches 92; Indels 0; Gaps 0;

```

QY      54  tatgttttaataaggaagaattttttcttcctttttatcacacctgacacacaga 113
        ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB       62  ttttttttgcaaaigcgaatttvtggaaakttttkttvkttktttttttagdka 143
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY      114  ctacagtcttagaacgccctaccgcgaataagaataaatccttgcaccttgtat 173
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB       142  ttcaaatttcaaaaaggagaaattcttcmtwaattscgtttkttwatttgcgaatktttt 83
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY      174  ttcaacagcaactctaattttttcttcctccanagctctcaaaaaaataaaaaa 233
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB       82  gtwcanaattactthtttttttttttttttttttttttttttttttttttttt 23
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY      234  accttactaaataacaggg 251
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB       22  AAAAAAAAAAAAAAAGGG 5
        ||||| | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 10
CONS165/c

LOCUS
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN15909 of BROSAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106426
VERSION AL106426.1 GI:5621382
KEYWORDS GSS.
SOURCE fruit fly,
ORGANISM Drosophila melanogaster
Fukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1201)

REFERENCE
AUTH-AS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National do Sequençage;
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC

```
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT      126 a      69 c      88 g      84 t
ORIGIN

Query Match      8.5%; Score 40.6; DB 10; Length 367;
Best Local Similarity 54.3%; Pred. No. 5.7e+02;
Matches 82; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 98 cacctgaccacaggaacttagcagttcttaggaagcccccttaccgcgaataaggaataaa 157
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87 CACATGCTCTGCGCTGCTCTCCCTCGTGTCTTACACCAACGATGAAGTAAA 146
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 158 tccttgcacattgatttcaaggccaatgctctaattttttcttctccagagctctcaa 217
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 147 GATTGTGACAGAGTTTTGAAAAAATAAACCCTTTCATGTGGCCCAAAAAA 206
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 218 aaaaaaataaaaccccttactataaaca 248
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 AAAAAAATAAAAAAATAAAAAAATAAAAAA 237
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
CNS004YY
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TPT3 end of BAC #
BACR11F03 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL055406
VERSION
AL055406.1 GI:4932207
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 859)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
JOURNAL
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
Location/Qualifiers
1..859
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR11F03"
/note="end : TPT3"
BASE COUNT      302 a      32 c      15 g      124 t      386 others
ORIGIN

Query Match      8.4%; Score 40.2; DB 12; Length 859;
Best Local Similarity 16.4%; Pred. No. 4.2e+02;
Matches 27; Conservative 85; Mismatches 53; Indels 0; Gaps 0;

```

```
Qy 74 ttttttctctccctttttgtaacacctgaccacagcagctgacagttctaggagccc 133
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 WHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHY
752
Qy 134 ccttaccgcgaataaggaataataatccttgcacacctgtattgtcgaaggccaatgctaatt 193
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 YYYWYTYAHYAWYAYYYYAAAAAYAYYYYAYYYYAYYYYAYYYYAYYYYAYYAY 812
Qy 194 ttttcttctccagagctctcaaaaaaataaaaaaaacacctt 238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 HAYAYYYYHYAYMYAYYAHYHYHAAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 857
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
CNS00GQG/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR33J07 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL072416
VERSION
AL072416.1 GI:4952297
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 980)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
JOURNAL
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
Location/Qualifiers
1..980
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR33J07"
/note="end : T7"
BASE COUNT      297 a      115 c      162 g      235 t      171 others
ORIGIN

Query Match      8.4%; Score 40.2; DB 12; Length 980;
Best Local Similarity 29.2%; Pred. No. 3.9e+02;
Matches 50; Conservative 56; Mismatches 65; Indels 0; Gaps 0;

Qy 75 ttttttctctccctttttgtaacacctgaccacagcagctgacagttctaggagcccc 134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 934 TTTTTHHHMMMMCMCTTTTSTMMMMMSAMTTCMMHHHHHHHTTMMMMMMMMCMCBTWT 875
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 135 cttaccgcgaataaggaataataatccttgcacacctgtattgtcgaaggccaatgctaatt 194
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 874 MCCHYCMCMCTCGSCMMCMCMCMTCMCCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 815
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 195 ttttcttctccagagctctcaaaaaaataaaaaaaacaccttactaaaaa 245
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

Search completed: July 7, 2002 00:43:14
 Job time: 6795 sec

[illegible]

